

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

Кваліфікаційна наукова
праця на правах рукопису

ЄФРЕМОВ КОСТЯНТИН ВІКТОРОВИЧ

УДК 004.75

ДИСЕРТАЦІЯ
ПЛАТФОРМА ПІДТРИМКИ МІЖДИСЦИПЛІНАРНИХ ДОСЛІДЖЕНЬ
В СВІТОВІЙ СИСТЕМІ ДАНИХ

Спеціальність 05.13.06 – інформаційні технології

Подається на здобуття наукового ступеня кандидата технічних наук

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Єфремов К.В.

Науковий керівник

Згуровський Михайло Захарович, доктор технічних наук, професор

Київ - 2021

АНОТАЦІЯ

Єфремов К.В. Платформа підтримки міждисциплінарних досліджень в Світовій системі даних. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 «Інформаційні технології». – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Міністерство освіти і науки України, Київ, 2021.

Інформаційні технології (ІТ) сприяють створенню нових і зручному доступу до накопичених інформаційних ресурсів, їх ефективному використанню завдяки консолідації, віртуалізації, впровадженню сервісного підходу. Одним із найпотужніших інформаційних ресурсів володіє Світова система даних (ССД). Накопичені в цій системі дані багатьох галузей науки є складовою потенціалу суспільства, яку можна використовувати для вирішення його актуальних проблем. Однак на перешкоді постають труднощі у взаємодії успадкованих застосунків, необхідність узгодження даних, отриманих від різних джерел, мінливість запитів, що вимагають широкого арсеналу засобів аналітичного опрацювання даних, швидкого створення нових застосунків, незручний доступ до інформаційних ресурсів для науковців, які не є фахівцями в галузі ІТ. Виникає потреба у платформі підтримки міждисциплінарних досліджень в Світовій системі даних, здатній усунути зазначені перешкоди.

Дисертаційна робота присвячена вирішенню важливої науково-практичної задачі організації, планування і здійснення підтримки міждисциплінарних досліджень в Світовій системі даних. Запропонована платформа підтримки міждисциплінарних досліджень в Світовій системі даних забезпечує інтеграцію застосунків і джерел даних, узгодження даних, впровадження проблемно-орієнтованого апарату інтелектуального аналізу даних та швидкого створення інформаційних ресурсів.

Створена платформа підтримки міждисциплінарних досліджень в

Світовій системі даних базується на розробленій в роботі концепції, яка об'єднує такі положення, підходи, моделі і методи: сучасне ІТ-середовище з дружнім доступом до успадкованої сукупності джерел даних і застосунків, яке інтегрує виокремлені системи збору, збереження і оброблення даних ССД на технологічному, змістовному і функціональному рівнях; взаємодія сервісів на основі оркестрації, що дозволяє групувати базові сервіси в ієрархічно скомплексовані системи, об'єднуючи їх за потрібними ознаками (галузь науки, функціональність та ін.); простий запит виконується одним сервісом, а складний запит, притаманний міждисциплінарним дослідженням, – композицією сервісів, яка будується сервісами-оркестраторами з використанням методів логічного виведення; запит користувача виконується на основі виводу, деревовидна структура якого дозволяє виконувати досить складні композиції сервісів використовуючи вказівники на потрібні джерела даних і сервіси; потрібні для цього залежності, пов'язані зі знаннями про функціональність сервісів і семантику даних, описуються в аксіомах, які вводяться в базу знань при реєстрації сервісів в системі, і правилах виведення, які визначають композицію сервісів; підтримка формування запитів в термінах предметної області, планування їх виконання і організації взаємодії базових сервісів в системі за допомогою інтелектуальних агентів, які реалізують логіку функціонування і взаємодії компонентів платформи; використання рішення класу ЕП з посередником і web-орієнтованою архітектурою для інтеграції джерел даних і створення складних запитів, що передбачають використання даних з різних джерел, існуючих застосунків і реалізації додаткових алгоритмів оброблення даних; використання UDDI для створення реєстру сервісів, WSDL – для уніфікованого їх описання, SOAP – для обміну повідомленнями, BPEL – для оркестрації сервісів.

Для реалізації концепції розроблено оригінальний комплекс моделей і алгоритмів інтеграції застосунків і джерел даних Світових центрів даних (СЦД), механізм міжагентної взаємодії, моделі і методи узгодження даних,

модель і метод генерування проблемно-орієнтованих застосунків на основі інтеграції мікросервісів.

Розвинена клаузальна логіка взаємодії застосунків дозволяє у зручному для користувача вигляді визначати потребу і оптимальним способом реалізувати відповідну схему взаємодії застосунків. Метод виведення для цієї логіки удосконалено за рахунок попереднього виведення в просторі типів з наступною деталізацією виводу в системі індивідних об'єктів.

Модифікована логіка описання джерел, механізми федерування, побудови запитів, об'єднання результатів для інтеграції джерел даних дозволяють користувачеві визначити потрібні дані, а системі спланувати, оптимізувати і виконати схему запитів до застосунків і джерел даних для отримання потрібних даних. Тим самим створена основа організації, планування і здійснення підтримки міждисциплінарних досліджень в Світовій системі даних. Завдяки цим моделям і методам програмуючий користувач може створювати ефективні схеми виконання складних запитів, що вимагають використання багатьох баз даних та існуючих застосунків, реалізації нових алгоритмів обчислень.

Оскільки використовувані в міждисциплінарних дослідженнях дані мають різну природу, визначену їх об'єктивним змістом, цільовим призначенням і способом отримання, в роботі розроблені моделі і методи системного узгодження даних різної природи, призначені для приведення даних до єдиної семантики, єдиного діапазону значень і єдиних одиниць виміру за умови мінімізації інформаційних втрат такого узгодження. Запропоновані оцінки інформаційних втрат і зростання невизначеності моделі, що супроводжують цей процес, методологія кількісного оцінювання узгодженості даних різної природи і методи узгодження перекривають потреби перетворення даних з контрольованими втратами при виконанні міждисциплінарних досліджень в ССД.

Моделі системного рівня організації мультиагентної системи (МАС) і структурного рівня організації агента втілюють ідею колективної діяльності

сукупності агентів. Для визначення семантики даних використовується Web Ontology Language (OWL). Мультиагентний підхід забезпечує придатність для подальшого розширення функціональності шляхом впровадження агентів, що забезпечують певні перетворення даних, інтеграцію даних відповідно до їх семантично однорідних фрагментів і візуалізацію у вигляді карт, графіків та інших представлень.

На основі концепції і запропонованого комплексу моделей, методів і алгоритмів розроблено інструментальні засоби платформи підтримки міждисциплінарних досліджень в ССД, яка забезпечує інтеграцію застосунків і джерел даних та планування і реалізацію визначених користувачем показників оцінювання процесів різноманітної природи на основі накопичених у Світовій системі даних ресурсів.

Основними архітектурними компонентами платформи є:

- 1) підсистема довготривалого зберігання даних, яка має функціональність, необхідну для вирішення задач управління даними;
- 2) підсистема опрацювання даних, яка реалізує алгоритми попереднього оброблення, верифікації та аналітичного опрацювання даних;
- 3) підсистема підтримки прикладних застосунків, що дозволяє вирішувати задачі розробки прикладних застосунків та управління ними;
- 4) загальний реєстр ресурсів, який дозволяє управляти конфігурацією розподіленої системи та здійснювати моніторинг її стану.

Розроблена платформа є основою розподіленої інформаційної системи з сервіс-орієнтованою архітектурою, призначеною для вирішення задач, що виникають в процесі реалізації безперервного життєвого циклу даних, які використовуються для наукових досліджень. Платформа має розвинуті механізми моніторингу стану індикаторів, аналізу даних та візуалізації. Набір засобів аналітичної обробки даних створено на основі сучасних ІТ, що забезпечує їх високу надійність та зручність у використанні. Кожен інструмент створено у вигляді окремого сервісу, що відповідає за роботу того чи іншого алгоритму оброблення даних.

Наукова новизна одержаних результатів визначається такими теоретичними і практичними результатами, отриманими автором:

1) вперше створено платформу підтримки міждисциплінарних досліджень в Світовій системі даних, яка забезпечує інтеграцію застосунків і джерел даних, оброблення запитів до розподілених джерел, системне узгодження даних різної природи, проблемно-орієнтоване інтелектуальне оброблення даних та автоматизоване генерування застосунків;

2) вперше створено метод узгодження даних, який відрізняється застосуванням оцінок інформаційних втрат процедур перетворення даних та оцінок узгодженості даних, отриманих з різних джерел, для конструювання процедури перетворення даних з найменшими інформаційними втратами;

3) модифіковано клаузальну логіку планування взаємодії застосунків для розв'язання проблем міждисциплінарних досліджень, яка відрізняється від відомих використанням системи типів, орієнтованих на організацію взаємодії застосунків, комплексом спеціальних аксіом описання застосунків на основі передумов та постумов і правил виведення, які визначають принципи комплексування застосунків, що дозволяє у зручному для користувача вигляді визначати проблему і будувати схему її вирішення.

4) отримав подальший розвиток метод виведення в клаузальній логіці взаємодії застосунків за рахунок попереднього виведення в просторі типів з наступною деталізацією виводу в системі індивідних об'єктів та відновленням схеми розв'язання проблеми на основі виводу.

Практичне значення платформи, а також розроблених моделей, методів, алгоритмів і інструментів полягає у створенні теоретичних основ розроблення, реалізації і впровадження інформаційних систем підтримки міждисциплінарних досліджень.

До результатів, які мають практичне значення, належать логічні моделі описання застосунків та джерел даних, визначення семантики предметної області, перетворення даних, агента та міжагентної взаємодії при розв'язанні проблем інтеграції даних та застосунків, модель і метод генерування

проблемно-орієнтованих застосунків на основі інтеграції мікросервісів, методи планування та підтримки співпраці агентів, методи та алгоритми виведення, відновлення дерева виконання запиту, федерування, побудови запитів, побудови плану виконання запиту із врахуванням коефіцієнтів ємності джерел даних, побудови зв'язаної структури даних із онтологій та підключень до джерел даних, об'єднання результатів, метод узгодження даних з врахуванням інформаційних втрат, методичні рекомендації щодо створення та налаштування проблемно-орієнтованих застосунків, які можуть бути використані в інформаційних системах, які потребують інтеграції джерел даних та застосунків, узгодження даних різної природи, проблемно-орієнтованого інтелектуального аналізу даних, швидкого розроблення і публікації інформаційних ресурсів.

Компоненти розробленої платформи підтримки міждисциплінарних досліджень в Світовій системі даних знайшли впровадження в Навчально-науковому комплексі «Світовий центр даних з геоінформатики та сталого розвитку» КПІ ім. Ігоря Сікорського, Корпорації «Науковий парк «Київська Політехніка», Інституті геофізики НАН України ім. С.І. Субботіна, Інституті проблем реєстрації інформації НАН України, Інформаційно-аналітичному ситуаційному центрі КПІ ім. Ігоря Сікорського та інших організаціях. Використання отриманих в дисертаційній роботі результатів підтвердило їх працездатність, дозволило, з одного боку, скоротити час проектування і реалізації застосунків, втрати внаслідок невчасного отримання потрібних даних, інформаційні втрати при перетворенні даних, з іншого боку, підвищити обсяги наданих користувачам інформаційних ресурсів, ефективність аналізу даних різної природи експертами без знання інструментальних мов програмування.

Ключові слова: Світова система даних, Світовий центр даних, інтеграція застосунків, інтеграція джерел даних, клаузальна логіка, системне узгодження даних, міждисциплінарні дослідження.

ABSTRACT

Kostiantyn Yefremov. A platform to support interdisciplinary research in the World Data System. – Manuscript of qualification scientific work.

Thesis for a candidate degree in technical sciences, specialty 05.13.06 – “Information Technologies”. – National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ministry of Education and Science of Ukraine, Kyiv, 2021.

Information technologies (IT) contribute to the creation of new and convenient access to accumulated information resources, their effective use through consolidation, virtualization, and implementation of a service approach. The World Data System (WDS) has one of the most powerful information resources. The data of many scientific fields accumulated in this system are a part of the society's potential, which can be used to solve its current problems. However, it is hampered by difficulties in the interaction of legacy applications, the need for data harmonization from different sources, the variability of queries that require a wide arsenal of analytical tools, the rapid creation of new applications, and inconvenient access to information resources for the non-IT-related scientists. There is a need for a platform to support interdisciplinary research in the World Data System that can address these barriers.

Dissertation work is devoted to solving the important scientific and practical problem of organizing, planning and implementing support for interdisciplinary research in the World Data System. The proposed platform to support interdisciplinary research in the World Data System provides integration of applications and data sources, data harmonization, implementation of problem-oriented data mining, and rapid creation of information resources.

The created platform to support interdisciplinary research in the World Data System is based on the concept developed in the dissertation, which combines the following findings, approaches, models, and methods: a modern IT environment

with friendly access to a legacy set of data sources and applications, which integrates separate systems for collecting, storing, and processing WDS data at the technological, conceptual, and functional levels; interaction of services on the basis of orchestration that allows to group basic services in hierarchically complex systems, combining them by the required attributes (scientific area, functionality, etc.); a simple query is performed by one service, and a complex query inherent in interdisciplinary research - by a composition of services, which is built by Orchestrator Services using the methods of inference; the user's query is performed upon formation of the proof, the tree structure of which allows performing quite complex compositions of services using pointers to the desired data sources and services; the dependencies required for this related to knowledge of service functionality and data semantics are described in the axioms that are added into the knowledge base when registering services in the system, and inference rules that determine the composition of services; support for the formation of queries in terms of the domain knowledge, planning their execution, and organization of basic services interaction in the system using intelligent agents that implement the logic of operation and interaction of platform components; the use of EII class solution with a mediator and web-oriented architecture for integration of data sources and construction of complex queries involving the use of data from different sources, existing applications, and the implementation of additional algorithms for data processing; the use of UDDI for creating a register of services, WSDL for a unified description of services; SOAP for exchanging notifications between services; BPEL – for services orchestration.

To implement the proposed concept the original set of models, methods, and algorithms for integration of World Data Centers' applications and data sources, mechanism of inter-agent interaction, models and methods of data harmonization, model and method of problem-oriented applications generation based on microservice integration have been developed.

Improved clausal logic of application interaction allows to determine the requirements in a user-friendly form and optimally implement the appropriate

scheme of application interaction. The inference method for this logic has been improved due to preliminary inference in the space of types with the following detailing inference in the system of individual objects.

Improved clausal logic of source description, mechanisms of the federation, query building, and combining of results for data sources integration allow the user to define the required data and allow the system to plan, optimize, and execute the query scheme to applications and data sources to obtain the required data. Thus established the basis for the organization, planning, and implementing support for interdisciplinary research in the World Data System. With these models and methods, the user familiar with programming can create efficient schemes for executing complex queries that require the use of many databases and existing applications, the implementation of new computational algorithms.

Since the data used in interdisciplinary research have a different nature, determined by their objective content, purpose, and generation method, the paper develops models and methods of system harmonization of data of various nature, designed to reduce these data to common semantics, a common range of values, and common units of measurement provided that information losses of such harmonization are minimized. The proposed estimates of information losses and estimates of the growing uncertainty of the model that accompany this process, the methodology for quantifying the consistency of data of various nature, and methods of data harmonization overlap the data transformation needs with controlled losses when performing interdisciplinary research in WDS.

Models of the system level of the multi-agent system (MAS) organization and the structural level of the agent's organization embody the idea of the collective actions of a set of agents. Web Ontology Language (OWL) is used to determine the semantics of the data. The multi-agent approach ensures suitability for further functionality enhancement by implementing agents that provide certain data transformations, data integration according to their semantically homogeneous fragments, and visualization in the form of maps, charts, and other representations.

The tools of the platform to support interdisciplinary research in the World Data System have been developed on the base of proposed concept and set of models, methods, and algorithms. The Platform provides integration of applications and data sources, processing queries to distributed sources, systemic harmonization of data of various nature, problem-oriented intelligent data processing and automated application generation.

The main architectural components of the Platform are:

- 1) Data Management Subsystem (long-term data storage and management)
- 2) Data Processing Subsystem (data pre-processing, data verification, and intelligent data processing)
- 3) Rapid Application Development Subsystem (application development, deployment, and management)
- 4) General Resource Registry (management of a distributed system configuration, status monitoring)

The developed Platform is the basis of a distributed information system with a service-oriented architecture, designed to solve problems that arise in the process of implementing a continuous life cycle of data used for scientific research. The Platform has developed mechanisms for monitoring the status of indicators, data analysis, and visualization. The set of analytical data processing tools is created based on modern IT, which provides their high reliability and ease of use. Each tool is created as a separate service that provides the operation of a particular data processing algorithm.

The scientific novelty of the obtained results is determined by the following theoretical and practical results obtained by the author:

- 1) for the first time a platform for supporting interdisciplinary research in the World Data System has been created. The Platform provides integration of applications and data sources, processing queries to distributed sources, systemic harmonization of data of various nature, problem-oriented intelligent data processing and automated application generation;

2) for the first time a method for the data harmonization has been developed, that differs by the use of estimates of information loss of data transformation procedures and estimates of consistency of data obtained from different sources, to design a procedure for converting data with the lowest information loss;

3) the clausal logic of planning the interaction of applications to solve problems of interdisciplinary research has been modified. It differs from the known by using the system of types, oriented to the interaction of applications, set of special axioms of application descriptions based on preconditions and postconditions, and inference rules, which define the application integration principles, which allows in a user-friendly form to identify the problem and to build the solution scheme;

4) the inference method in the clausal logic of application interactions has been further improved due to preliminary inference in the space of types with the following detailing inference in the system of individual objects and solution scheme reconstruction based on the proof generated by the inference mechanism.

The practical value of the Platform, as well as the developed models, methods, algorithms, and tools is to create theoretical fundamentals for the development, deployment, and implementation of information systems to support interdisciplinary research.

The most practical results include: logical models for describing applications and data sources, determining the semantics of the domain knowledge, data transformation, agent and inter-agent interaction in solving problems of data and application integration, model and method of generating problem-oriented applications based on the integration of microservices, methods of planning and supporting the cooperation of agents, methods and algorithms of inference and the restoring of the decision tree, federation, query building, building a query execution plan, taking into account factors capacity data sources, methods of building a linked data structure from ontologies and connections to data sources, combining results; method of the data harmonization, taking into account estimates of information loss of data transformation procedures, the methodological recommendations for creating and configuring problem-oriented applications that can be used in

information systems that require the integration of data sources and applications, harmonization of data of different nature, problem-oriented data mining, rapid development and publication of information resources.

Components of the developed platform to support interdisciplinary research in the World Data System were implemented at the Institute for Information Recording of the National Academy of Sciences of Ukraine, the Institute of Geophysics by S.I. Subbotin name of the National Academy of Sciences of Ukraine, the Information and Analytical Situation Center of Igor Sikorsky Kyiv Polytechnic Institute, the Educational and Scientific Complex "World Data Center for Geoinformatics and Sustainable Development" of Igor Sikorsky Kyiv Polytechnic Institute, Corporation "Science Park "Kyivska Polytechnika", and other organizations, which allowed to confirm the efficiency of the Platform, to reduce the time for applications design and implementation, to reduce losses due to late receipt of the required data, to reduce information losses in data transformation; to increase the amount of information resources provided to users, and to increase the efficiency of data analysis of various nature by experts without knowledge of instrumental programming languages.

Keywords: World Data System, World Data Center, application integration, data source integration, clausal logic, systemic data harmonization, interdisciplinary research.

Список публікацій здобувача, в яких опубліковано основні наукові результати дисертації:

1. Zgurovsky M., Boldak A., Melnyk O., Perestyuk M., Putrenko V., Pyshnograiev I., Yasinsky V., Yefremov K., Foresight 2018: systemic world conflicts and global forecast for XXI century / International Council for Science etc.; Scientific Supervisor M. Zgurovsky. – К. : NTUU «Igor Sikorsky Kyiv Polytechnic Institute», 2018. — 226 p. (колективна монографія).
2. Згуровський М.З., Болдак А.О., Джигирей І.М., Єфремов К.В., Форсайт 2018: Аналіз підготовки і перепідготовки фахівців природничого і технічного спрямування, виходячи з цілей сталого соціально-економічного розвитку України до 2025 року. — К. : НТУУ «КПІ імені Ігоря Сікорського», Вид-во «Політехніка», 2018. — 32 с. (колективна монографія).
3. Foresight and construction of the strategies of socio-economic development of Ukraine on mid-term (up to 2020) and long-term (up to 2030) time horizons / Scientific advisor of the project acad. of NAS of Ukraine M. Zgurovsky // International Council for Science (ICSU); Committee for the System Analysis of the Presidium of NAS of Ukraine; National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»; Institute for Applied System Analysis of MES of Ukraine and NAS of Ukraine; World Data Center for Geoinformatics and Sustainable Development; Agrarian Superstate Foundation. — 2nd ed. — Kyiv : NTUU «Igor Sikorsky KPI», Publ. house «Polytechnica», 2016. — 184 p. (колективна монографія).
4. Sustainable Development Analysis: Global and Regional Contexts / International Council for Science (ICSU) and others; Scientific Supervisor of the Project M. Zgurovsky. — К. : Igor Sikorsky KPI, 2017. — Part 2. Ukraine in Sustainable Development Indicators (2016–2017). — 72 p. (колективна монографія).
5. Sustainable Development Analysis: Global and Regional Contexts / International Council for Science (ICSU) and others; Scientific Supervisor of the

- Project M. Zgurovsky. — K. : Igor Sikorsky KPI, 2017. — Part 1. Global Analysis of Quality and Security of Life (2016). — 208 p. (колективна монографія).
6. Marsel Shaimardanov, Alexei Gvishiani, Michael Zgurovsky, Alexander Sterin, Alexander Kuznetsov, Natalia Sergeyeva, Evgeny Kharin, Kostiantyn Yefremov. Development of WDS Russian-Ukrainian segment // Data Science Journal. — 2013. — Volume 12. — p. 17-26. (журнали країн, що входять до ОЕСР, ЄС; включене до бази даних Scopus).
 7. M.Z. Zgurovsky, A.A. Boldak, K.V. Yefremov. Intelligent analysis and the systemic adjustment of scientific data in interdisciplinary research // Cybernetics and Systems Analysis: Volume 49, Issue 4 (2013), pp. 541-552. (журнал категорії «А», включене до бази даних Scopus).
 8. Теленик С.Ф., Амонс О.А., Єфремов К.В., Лиско В.Т. Логічний підхід до інтеграції програмних застосувань підтримки міждисциплінарних наукових досліджень // Наукові вісті НТУУ «КПІ». — 2013. — № 5. — С. 53-72.
 9. Теленик С.Ф., Амонс О.А., Єфремов К.В., Жук С.В. Семантична інтеграція різномірних інформаційних ресурсів // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: Зб. наук. пр. — К.: Век+, — 2013. — No. 58. — с.29-45.
 10. Згуровский М.З., Болдак А.А., Ефремов К.В., Сергеева Н.А., Забаринская Л.П., Шестопапов И.П., Нисилевич М.В. Применение методов интеллектуального анализа данных для эмпирических исследований взаимосвязи гелио- и геофизических процессов // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: збірник наукових праць. — К.: Век+, — 2013. — No. 58. — с.4-10.
 11. Болдак А.О., Єфремов К.В. Предметно-орієнтована мова аналітичної обробки даних // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: збірник наукових праць. — К.: Век+, — 2012. — No. 55. — с.67-71.

12. M. Z. Zgurovsky, A. D. Gvishiani, K. V. Yefremov and A. M. Pasichny. Integration of the Ukrainian science into the world data system // *Cybernetics and Systems Analysis: Volume 46, Issue 2 (2010)*, pp. 211-219. (журнал категорії «А», включене до бази даних Scopus).
13. M. Zgurovsky, A. Boldak, D. Lande, K. Yefremov and M. Perestyuk. Predictive Online Analysis of Social Transformations based on the Assessment of Dissimilarities between Government Actions and Society's Expectations. Conference proceedings of 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2020, P. 130-135. (включене до бази даних Scopus).
14. Nowakowski, G., Telenyk, S., Yefremov, K., Khmeliuk, V. The approach to applications integration for world data center interdisciplinary scientific investigations. Proceedings of the 2019 Federated Conference on Computer Science and Information Systems. FedCSIS 2019, P. 539-545. (включене до бази даних Scopus).
15. Zgurovsky, M., Putrenko, V., Dzhygyrey, I., Boldak, A., Yefremov, K., Pashynska, N., Pyshnograiev, I., Nazarenko, S. Parameterization of Sustainable Development Components Using Nightlight Indicators in Ukraine. Conference proceedings of 2018 IEEE 1st International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2018, P. 1-5. (включене до бази даних Scopus).
16. Zgurovsky, M., Boldak, A., Yefremov, K., Pyshnograiev, I. Modeling and investigating the behavior of complex socio-economic systems. 2017. IEEE 1st Ukraine Conference on Electrical and Computer Engineering, UKRCON 2017 – Proceedings, P. 1113-1116. (включене до бази даних Scopus).
17. Telenyk, S., Nowakowski, G., Yefremov, K., Khmeliuk, V. Logics based application integration for interdisciplinary scientific investigations. Proceedings of the 2017 IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2017, P. 1026-1031. (включене до бази даних Scopus).

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	19
ВСТУП	21
РОЗДІЛ 1. АНАЛІЗ ПРОБЛЕМ ОРГАНІЗАЦІЇ МІЖДИСЦИПЛІНАРНИХ ДОСЛІДЖЕНЬ В СВІТОВІЙ СИСТЕМІ ДАНИХ	34
1.1. Міждисциплінарні дослідження: проблеми і перспективи	34
1.2. Сутність та особливості організації і підтримки міждисциплінарних досліджень в Світовій системі даних	42
1.3. Проблеми підтримки міждисциплінарних досліджень в Світовій системі даних	48
1.4. Огляд існуючих рішень	59
1.5 Уточнення задачі дослідження	75
Висновки за розділом	79
РОЗДІЛ 2. МОДЕЛІ І МЕТОДИ ІНТЕГРАЦІЇ ЗАСТОСУНКІВ І ДЖЕРЕЛ ДАНИХ	81
2.1. Концепція інтеграції застосунків і джерел даних в рамках платформи підтримки міждисциплінарних досліджень в Світовій системі даних	81
2.2. Логічна модель як формальна основа інтеграції застосунків	96
2.3 Метод виведення в клаузальній логіці взаємодії застосунків	103
2.4. Механізм відновлення схеми вирішення проблеми на основі виводу	108
2.5. Приклад застосування логічного підходу до вирішення задач сталого розвитку регіонів України у СІД-Україна	110
Висновки за розділом	118

РОЗДІЛ 3. МОДЕЛІ І МЕТОДИ ІНТЕГРАЦІЇ ДЖЕРЕЛ ДАНИХ І СИСТЕМНОГО УЗГОДЖЕННЯ ДАНИХ РІЗНОЇ ПРИРОДИ	120
3.1. Моделі і методи інтеграції джерел даних в рамках платформи підтримки міждисциплінарних досліджень в Світовій системі даних.....	120
3.2. Моделі міжагентної взаємодії.....	135
3.3. Системне узгодження даних різної природи	144
Висновки за розділом	155
РОЗДІЛ 4. РОЗРОБЛЕННЯ ПЛАТФОРМИ ПІДТРИМКИ МІЖДИСЦИПЛІНАРНИХ ДОСЛІДЖЕНЬ І ЇЇ ЗАСТОСУВАННЯ В СВІТОВІЙ СИСТЕМІ ДАНИХ	156
4.1. Архітектура платформи і принципи її функціонування	156
4.2. Реалізація платформи підтримки міждисциплінарних досліджень	166
4.3. Застосування платформи для міждисциплінарних досліджень на основі моделей і методів інтелектуального аналізу даних із різних джерел	176
Висновки за розділом	193
ВИСНОВКИ	194
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	196
Додаток А. Приклад моделювання процесів сталого розвитку на платформі міждисциплінарних досліджень СІЦД-Україна	212
Додаток Б. Оцінювання ставлення суспільства до проведення в Україні масової вакцинації проти COVID-19, виконане на платформі міждисциплінарних досліджень СІЦД-Україна	221
Додаток В. Список публікацій здобувача за темою дисертації, відомості про апробацію результатів та документи, які підтверджують результати впровадження	231

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

AIC	Автоматизовані інформаційні системи
АОП	Агентно-орієнтоване проектування
БД	База даних
ДКР	Дослідно-конструкторська робота
ІКТ	Інформаційно-комунікаційні технології
ІС	Інформаційна система
ІТ	Інформаційна технологія
ІТС	Інформаційно-телекомунікаційна система
МАС	Мультиагентна система
МНР	Міжнародна наукова рада
НДР	Науково-дослідна робота
ПЗ	Програмне забезпечення
РК	Реєстраційна картка
СЗД	Система збереження даних
СОА	Сервіс-орієнтована архітектура
ССД	Світова система даних
СУБД	Система управління базами даних
СЦД	Світовий центр даних
ЦОД	Центр оброблення даних
ANSI	American National Standards Institute
BPEL	Business Process Execution Language
BSUN	Black Sea Universities Network
CORBA	Common Object Request Broker Architecture
CSV	Comma-Separated Values
DAML	DARPA Agent Markup Language
DARPA	Defense Advanced Research Projects Agency
EAI	Enterprise Application Integration
EII	Enterprise Information Integration
ETL	Extract, Transform, Load

FIPA	Foundation for Intelligent Physical Agents
GCMD	Global Change Master Directory
HTML	HyperText Markup Language
IDL	Interface Definition Language
IEEE	Institute of Electrical and Electronics Engineers
JADE	Java Agent DEvelopment Framework
JDBC	Java DataBase Connectivity
OIL	Ontology Interchange Language
OML	Ontology Markup Language
OWL	Ontology Web Language
P2P	Peer-to-peer
RDA	Research Data Alliance
RDF	Resource Description Framework
RDFS	RDF Schema
REST	Representational state transfer
SGML	Standard General Markup Language
SOAP	Simple Object Access Protocol
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Structured Query Language
STCU	Science & Technology Center in Ukraine
UDDI	Universal Description Discovery & Integration
URI	Uniform Resource Identifier
UUID	Universally Unique Identifier
W3C	World Wide Web Consortium
WCF	Windows Communication Foundation
WDS	World Data System
WSDL	Web Services Description Language
XML	eXtensible Markup Language

ВСТУП

Актуальність теми. Сьогодні інформаційна система (ІС) і інформаційна технологія (ІТ) становлять незмінну складову професійного середовища у всіх сферах діяльності людини. Вони значною мірою сприяють науково-технічному прогресу, оскільки підтримують власне наукову діяльність, збільшуючи продуктивність праці науковців, розширюючи інструментарій дослідження, створюючи умови для ефективного використання обчислювальних, інформаційних та інших ресурсів.

Консолідація ресурсів, їх віртуалізація, сервісний підхід приводять до заміни локальних рішень розподіленими, які дозволяють комплексно використовувати всі об'єднані в глобальну мережу обчислювальні потужності, системи збереження даних і забезпечують доступ до накопичених інформаційних ресурсів. Але раціонально використовувати нові можливості можуть лише фахівці в галузі ІТ, які власне і створюють сучасні ІС та ІТ. Однак сучасні ІС та ІТ будуються таким чином, щоб за рахунок зручного інтерфейсу, використання орієнтованих на опис предметної області мов запитів, орієнтованих на реалізацію процесів певної визначеної природи, спеціалізованих мов програмування надати можливість раціонально використовувати ІС та ІТ фахівцям, які досконало знають свою предметну область [1].

З одного боку, це є проявом загальної тенденції до формування нового демократичного ІТ-середовища, в якому надавати сервіси можуть навіть невеликі підприємства. З іншого боку, природне розширення арсеналу засобів розроблення ІС та ІТ, урізноманітнення засобів і технологій доступу до інформаційних ресурсів супроводжується вимогами забезпечення уніфікованого доступу до сервісів, щоб ефективно використовувати переваги розподілених систем і можливості сервіс-орієнтованих технологій.

Так, одним із потужних інформаційних ресурсів користується створена в 1956 р. під егідою Міжнародної наукової ради (International Science Council – ISC, <https://council.science/>) Світова система даних (World Data System –

WDS, CCD, <https://www.worlddatasystem.org/>), яка забезпечує збір, збереження, поширення і аналіз даних, отриманих в багатьох галузях науки [2]. У CCD накопичено багато даних і застосунків, які можна використовувати для вирішення актуальних проблем розвитку суспільства. Але цьому перешкоджають труднощі у взаємодії успадкованих застосунків, використанні даних із різних джерел, формуванні результатів опрацювання запитів, що вимагають аналітичного опрацювання даних із різних джерел, застосування наявних і створення нових застосунків, обумовлені відмінністю архітектур, різноманітням форматів представлення даних та іншими чинниками.

Виникає потреба в новому підході до організації міждисциплінарних досліджень у Світовій системі даних, здатній забезпечити інтеграцію застосунків і ресурсів із різних джерел, виконання комплексного аналітичного оброблення даних. Саме при такому підході вчені різних галузей знання зможуть використовувати накопичені ресурси, орієнтуючись на знання предметної області, а не особливості ІТ.

До кола джерел, інтегрований доступ до ресурсів яких доцільно забезпечити, належать бази даних, сайти і портали, різноманітні успадковані файлові системи і репозиторії структурованих за різними моделями даних. Зазвичай інтеграцію даних в ІС розуміють як забезпечення єдиного уніфікованого інтерфейсу для доступу до певної сукупності, у загальному випадку, неоднорідних незалежних джерел даних. Для користувача інформаційні ресурси всієї сукупності джерел, які інтегруються, повинні подаватися як нове єдине джерело, а система, яка забезпечує такі можливості, називається системою інтеграції даних.

Погляд на інтеграцію як взаємодію застосунків не такий поширений, але дуже важливий з точки зору вирішення комплексної проблеми інтеграції в новому ІТ-середовищі. Загалом він формувався у середовищі фахівців з автоматизації програмування, але поступово через проблематику інтеграції мультібаз і федеративних баз даних, великих сховищ даних, різноманітних репозиторіїв був перенесений на інтеграцію web-застосунків. Цей напрям

сьогодні видається надзвичайно важливим, оскільки в умовах сервісного підходу його можна розглядати як теоретико-методологічну основу впровадження компонентно-базованого підходу до створення ІС.

Створення систем аналітичного оброблення даних, орієнтованих на користувача-непрофесіонала в ІТ, пов'язане зі створенням нових мов запитів, орієнтованих на семантичне представлення інформації, автоматизацію програмування і розроблення комплексних рішень для аналітичного оброблення міждисциплінарних даних.

Тому реалізація нового підходу до організації міждисциплінарних досліджень у ССД передбачає впровадження комплексного рішення, яке забезпечує інтеграцію даних в інформаційних системах і інтеграцію як взаємодію застосунків, швидке використання наявних і створення нових застосунків, широкий вибір засобів аналізу міждисциплінарних даних. На сучасному етапі розвитку ІТ-галузі мова йде про розроблення платформи підтримки міждисциплінарних досліджень у Світовій системі даних, яка дозволить:

- інтегрувати застосунки шляхом створення механізмів ефективного управління інформаційними ресурсами;
- інтегрувати джерела даних шляхом розширення сфери стандартизації протоколів і форматів обміну даними на основі взаємодії відкритих систем;
- генерувати і розгортати проблемно-орієнтовані крос-платформні застосунки для аналітичного оброблення міждисциплінарних даних шляхом використання шаблонів і метамов описання інформаційних ресурсів СЦД;
- узгоджувати і оброблювати накопичені в СЦД дані за рахунок створення моделей системного узгодження даних і проблемно-орієнтованого апарату інтелектуального оброблення даних.

Вагомий внесок у вирішення цієї проблеми зробили такі відомі вітчизняні та зарубіжні вчені, як Михайло Згуровський, Олександр Павлов,

Євген Бодянський, Олексій Новіков, Петро Бідюк, Анатолій Каргін, Микола Глибовець, Леонід Гуляницький, Наталя Шаховська, Сергій Кривий, Володимир Опанасенко, Сергій Погорілий, Олександр Ролік, Трейсі Кімбрел, Олександр Клемм, Верн Паксон та ін. Дослідженнями, пов'язаними зі створенням такої платформи також займаються низка міжнародних організацій (WDS, ISC, CODATA, Research Data Alliance), провідні університети (Гарвард, Кембрідж, Кайзерслаутерн, Берклі та ін.). Однак, не зважаючи на значні досягнення, відсутні єдине системне бачення усіх аспектів проблеми, сутності та призначення інструментарію міждисциплінарних досліджень і його місця в ССД, цілісна концепція платформи підтримки міждисциплінарних досліджень з урахуванням всього комплексу взаємозв'язків і взаємовпливів пов'язаних з нею процесів. Різний ступінь глибини опрацювання окремих аспектів цієї проблеми обумовив потребу в ефективних моделях і методах інтеграції джерел і застосунків, аналітичного оброблення міждисциплінарних даних, узгодження даних, швидкого розроблення застосунків.

Важливість проблеми підкреслює її міждисциплінарний характер. Сучасні наукові дослідження, узгоджені з пошуком відповідей на глобальні виклики, що виникають на початку XXI століття, є міждисциплінарними та спрямованими на вирішення погано структурованих завдань. Прикладом таких досліджень є аналіз процесів сталого розвитку в глобальному та регіональному контексті.

Міждисциплінарні дослідженнями передбачають використання кількісних або якісних оцінок, що характеризують різні явища або об'єкти. Для узгодження цих даних і моделей, розроблених у різних наукових сферах, необхідно розробити узагальнені (міждисциплінарні) моделі повної презентації об'єкта дослідження. Відсутність формальних моделей, використання результатів об'єктивних вимірювань разом з суб'єктивними експертними оцінками визначають погану структурованість завдань такого дослідження і додають важливі аспекти до його актуальності.

Вирішення проблеми базується на методах наукових розрахунків, які визначають форму і сутність концепцій інтелектуального оброблення даних. Інтелектуальне оброблення даних вимагає спеціальних програмних засобів та інформаційно-комунікативної інфраструктури, оскільки потребує використання великих обсягів даних різної природи. Необхідність витягнення з «сирих» даних невідомих, нетривіальних, практичних і корисних знань, які можуть бути інтерпретовані і використовувані для прийняття рішень в різних сферах людської діяльності, надає актуальності цієї проблеми сучасне трактування.

Необхідність гармонізації заходів щодо створення такої інфраструктури та організації єдиного інформаційного простору для підтримки вирішення фундаментальних і прикладних міждисциплінарних проблем, її впровадження і розвитку на прикладі функціонування одного СЦД і ССД в цілому визначають актуальну науково-практичну задачу організації, планування і здійснення підтримки міждисциплінарних досліджень в Світовій системі даних.

Вирішення цієї задачі пов'язане зі створенням платформи підтримки міждисциплінарних досліджень в Світовій системі даних і експериментальним дослідження її застосування на прикладі українського сегменту ССД. Це вимагає розроблення та дослідження принципів архітектурної організації відкритої платформи, на основі яких можна спроектувати, створити та інтегрувати її компоненти. Платформа надасть користувачам системи СЦД дані та інструменти для вирішення завдань міждисциплінарних досліджень.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційне дослідження виконано в рамках наукової тематики кафедри математичних методів системного аналізу Інституту прикладного системного аналізу та Навчально-наукового комплексу «Світовий центр даних з геоінформатики та сталого розвитку» КПІ ім. Ігоря Сікорського. Робота виконувалась в рамках таких держбюджетних і договірних науково-дослідних робіт (НДР) та проектів:

1) НДР «Розвиток комплексу баз даних та алгоритмів обробки з метою системного прогнозування поведінки складних антропогенних та природних систем» (2008-2009 рр., РК №0108U002348, НАН України);

2) НДР «Розвиток мережі Світових центрів даних для дослідження фундаментальних основ глобального моделювання складних природних та антропогенних систем» (2009-2010 рр., РК №0109U007160, НАН України);

3) НДР «Розробка фундаментальних основ та методів аналізу міждисциплінарних даних для створення системи інтегрованого доступу до інформаційних ресурсів Світових центрів даних Росії та України» (2010-2011 рр., РК №0110U003744, НАН України);

4) НДР «Розробка загального підходу та методів системного узгодження даних різної природи в інфраструктурі розподілених багатодисциплінарних баз даних Російсько-Українського сегменту Світової Системи Даних для вирішення фундаментальних міждисциплінарних задач взаємозв'язку процесів в системі геосфер» (2012-2013 рр., РК №0113U005058, НАН України);

5) НДР «Розробка методології та математичних основ кількісного оцінювання процесів сталого розвитку та впливу на них сукупності глобальних загроз у глобальному та регіональному контекстах» (2011-2013 рр., РК №0111U001738, МОН України);

6) НДР «Розробка інструментарію для побудови складних багатокомпонентних моделей» (2012-2013 рр., РК №0112U000557, МОН України);

7) НДР «Розробка інструментарію для аналізу сталого розвитку муніципалітетів» (2014-2015 рр., РК №0114U000671, НАН України);

8) НДР «Методологія та інструментарій стратегічної екологічної оцінки проектів розвитку регіонів України», 2015-2016 рр., РК №0115U001968, РК №0116U006141, ДФФД України);

9) НДР на замовлення підприємств і організацій України «Розробка та впровадження інформаційно-технічних засобів для збирання, обміну та

спільного використання даних в галузі зняття з експлуатації атомних енергоблоків, поводження з радіоактивними відходами, ядерної та радіаційної безпеки, перетворення об'єкта «Укриття» в екологічно безпечну систему, реабілітації забруднених територій» (2016 р., ДСП «Чорнобильська АЕС»);

10) Проєкт «Розробка інформаційних технологій для моделювання, кількісного оцінювання та прогнозування впливу загроз виникнення конфліктів та розповсюдження зброї на стабільність розвитку суспільства в регіональному та глобальному масштабах» (2016-2017 рр, проєкт №6166, Український науково-технологічний центр (STCU);

11) НДР на замовлення підприємств і організацій України «Форсайт та побудова стратегії соціально-економічного розвитку України на середньостроковому (до 2020 року) і довгостроковому (до 2030 року) часових горизонтах» (2016 р., Кооператив «Перший національний аграрний кооператив», ГО «Фундація Аграрна наддержава»);

12) НДР «Розробка он-лайн платформи аналізу і сценарного планування сталого розвитку регіонів України в контексті якості та безпеки життя людей» (2017-2019 рр., РК №0117U002476, МОН України);

13) НДР «Інтегрована платформа для оцінювання та сценарного планування сталого розвитку об'єднаних територіальних громад в ході проведення адміністративно-територіальної реформи в Україні» (2020-2022 рр., РК 0120U102060, МОН України).

Мета і завдання дослідження. Метою роботи є підвищення ефективності використання інформаційних та програмних ресурсів Світової системи даних за рахунок розроблення моделей, методів та інструментів інтеграції застосунків і джерел даних, узгодження та інтелектуального оброблення даних різної природи та створення на їх основі платформи підтримки міждисциплінарних досліджень.

Для досягнення мети необхідно сформулювати єдине бачення процесів.

Враховуючи світовий досвід, загальні тенденції, регіональні та галузеві особливості Світової системи даних для досягнення мети у відповідності з

логікою дисертаційного дослідження, яка визначає послідовність, взаємозв'язок, взаємовплив і обґрунтованість процедур, пов'язаних із вирішенням задач інтеграції застосунків і джерел даних, у роботі були поставлені такі завдання:

- 1) дослідження проблеми інтеграції застосунків і джерел даних ССД;
- 2) розроблення концепції платформи підтримки міждисциплінарних досліджень в ССД;
- 3) розроблення комплексу моделей і методів інтеграції застосунків Світової системи даних;
- 4) розроблення комплексу моделей і методів інтеграції джерел даних Світової системи даних;
- 5) розроблення механізму генерування проблемно-орієнтованих застосунків для аналітичного оброблення міждисциплінарних даних;
- 6) розроблення комплексу моделей і методів системного узгодження даних і впровадження проблемно-орієнтованого апарату інтелектуального оброблення даних;
- 7) розроблення інструментальних засобів платформи підтримки міждисциплінарних досліджень в ССД;
- 8) проведення експериментальних досліджень платформи підтримки міждисциплінарних досліджень в ССД.

Об'єкт дослідження дисертаційної роботи – процеси підтримки міждисциплінарних досліджень Світової системи даних.

Предмет дослідження – математичні моделі, методи та інструменти інтеграції застосунків і джерел даних, а також методи узгодження та інтелектуального оброблення даних різної природи.

Методи досліджень, застосовані для досягнення мети і вирішення поставлених завдань:

- 1) методи загальної теорії систем, теорії множин, формально-аналітичного моделювання – для аналізу діяльності, інформаційних ресурсів

та інфраструктури центрів Світової системи даних як об'єкта управління, розроблення концепції платформи підтримки міждисциплінарних досліджень в ССД;

2) методи математичного програмування, дискретної математики та теорії штучного інтелекту – для розроблення методів інтеграції застосунків і джерел даних Світової системи даних;

3) методи теорії імовірності і математичної статистики, системного аналізу – для оцінювання показників соціальних систем, країн, регіонів, об'єктів різноманітного призначення, що є предметом запитів користувачів Світової системи даних;

4) моделі та методи виведення математичної логіки – для планування, виконання запитів до застосунків і джерел даних Світової системи даних.

Наукове значення роботи полягає у вирішенні важливої науково-практичної задачі розроблення моделей, методів і комплексного інструментарію підтримки міждисциплінарних досліджень Світової системи даних (інтеграції застосунків і джерел даних, узгодження даних, планування, формулювання, виконання запитів, об'єднання результатів, проблемно-орієнтованого інтелектуального аналізу даних, швидкого розроблення застосунків).

Наукова новизна одержаних результатів полягає у наступному:

- *вперше* створено платформу підтримки міждисциплінарних досліджень в Світовій системі даних, яка забезпечує інтеграцію застосунків і джерел даних, оброблення запитів до розподілених джерел, системне узгодження даних різної природи, проблемно-орієнтоване інтелектуальне оброблення даних та автоматизоване генерування застосунків;
- *вперше* створено метод узгодження даних, який відрізняється застосуванням оцінок інформаційних втрат процедур перетворення даних та оцінок узгодженості даних, отриманих з різних джерел, для

конструювання процедури перетворення даних з найменшими інформаційними втратами;

- *модифіковано* клаузальну логіку планування взаємодії застосунків для розв'язання проблем міждисциплінарних досліджень, яка відрізняється від відомих використанням системи типів, орієнтованих на організацію взаємодії застосунків, комплексом спеціальних аксіом описання застосунків на основі передумов та постулов і правил виведення, які визначають принципи комплексування застосунків, що дозволяє у зручному для користувача вигляді визначати проблему і будувати схему її вирішення.
- *отримав подальший розвиток* метод виведення в клаузальній логіці взаємодії застосунків за рахунок попереднього виведення в просторі типів з наступною деталізацією виводу в системі індивідних об'єктів та відновленням схеми розв'язання проблеми на основі виводу.

Практичне значення отриманих результатів полягає в тому, що запропонована концепція, моделі, методи, методики і платформа підтримки міждисциплінарних досліджень в Світовій системі даних складають основу для створення інформаційних систем підтримки міждисциплінарних досліджень Світової системи даних. Практичне значення мають логічні моделі опису застосунків та джерел даних, визначення семантики предметної області, перетворення даних, агента та міжагентної взаємодії при розв'язанні проблем інтеграції даних та застосунків, модель і метод генерування проблемно-орієнтованих застосунків на основі інтеграції мікросервісів, методи планування та підтримки співпраці агентів, методи та алгоритми виведення, відновлення дерева виконання запиту, федерування, побудови запитів, побудови плану виконання запиту із врахуванням коефіцієнтів ємності джерел даних, побудови зв'язаної структури даних із онтологій та підключень до джерел даних, об'єднання результатів, метод узгодження даних з врахуванням

інформаційних втрат, методичні рекомендації щодо створення та налаштування проблемно-орієнтованих застосунків.

На їх основі можна проектувати, реалізувати і впроваджувати автоматизовані інформаційні системи (AIC) для менеджерів і користувачів різних рівнів Світових центрів даних, розвивати їх компоненти у відповідності зі змінами умов їх функціонування.

Застосування реалізованої на основі розроблених моделей, методів, методик платформи підтримки міждисциплінарних досліджень забезпечує підвищення ефективності використання інформаційних та програмних ресурсів Світової системи даних Міжнародної наукової ради, скорочення витрат на створення, впровадження і підтримку функціонування проблемно-орієнтованих застосунків, підвищення доступності важливих ресурсів Світової системи даних, скорочення витрат на створення, впровадження і підтримку функціонування AIC різних рівнів Світових центрів даних.

Результати дисертаційної роботи використано в Світовому центрі даних з геоінформатики та сталого розвитку (Київ, Україна) при створенні інтегрованої платформи «Advanced Analytics» для ситуаційного аналізу соціально-економічних і безпекових явищ, на базі якої створено програмно-апаратний комплекс Інформаційно-аналітичного ситуаційного центру КПІ ім. Ігоря Сікорського.

Результати дисертаційної роботи також впроваджено в Інституті проблем реєстрації інформації НАН України, Інституті геофізики НАН України ім. С.І. Субботіна, Інформаційно-аналітичному ситуаційному центрі КПІ ім. Ігоря Сікорського, Навчально-науковому комплексі «Світовий центр даних з геоінформатики та сталого розвитку» КПІ ім. Ігоря Сікорського, Корпорації «Науковий парк «Київська Політехніка». Розроблені методичні та програмні засоби використовуються в навчальному процесі на кафедрі математичних методів системного аналізу та кафедрі автоматики та управління в технічних системах КПІ ім. Ігоря Сікорського при підготовці фахівців в галузі інтелектуального аналізу даних та інформаційних технологій.

Особистий внесок здобувача. Наукові положення, висновки, рекомендації, які викладені в дисертаційній роботі і виносяться на захист, розроблені особисто здобувачем. З праць, опублікованих у співавторстві, в роботі використані лише ті ідеї, положення і висновки, які є результатом особистої роботи автора і становлять його індивідуальний внесок. До основних результатів дисертаційного дослідження, отриманих автором, належать: логічні моделі опису застосунків та джерел даних [3, 4]; моделі визначення семантики предметної області [5]; модель перетворення даних [6]; модель і метод генерування проблемно-орієнтованих застосунків на основі інтеграції мікросервісів [7, 8]; методи планування та підтримки співпраці агентів [4, 9]; методи та алгоритми виведення, відновлення дерева виконання запиту [3-5, 10]; методи федерування, побудови запитів, побудови планів виконання запитів із врахуванням коефіцієнтів ємності джерел даних, об'єднання результатів [5]; архітектура системи інтеграції застосунків [10] та підхід до інтеграції сервісів програмних систем в Світовій системі даних для підтримки міждисциплінарних досліджень [11]; метод узгодження даних з врахуванням інформаційних втрат [6]; алгоритм побудови зв'язаної структури даних із онтологій та підключень до джерел даних [5]; методичні рекомендації щодо створення та налаштування проблемно-орієнтованих застосунків [8, 12, 13].

Автором розроблено платформу підтримки міждисциплінарних досліджень в Світовій системі даних та реалізовано засобами платформи тематичні застосунки, мікросервіси, віджети та сценарії (сукупність взаємопов'язаних процесів) обробки даних для вирішення низки прикладних задач [8, 12-15, 16-19].

Апробація результатів дисертації. Основні положення дисертації розглядалися і обговорювалися на спільному засіданні кафедри математичних методів системного аналізу, кафедри автоматизації та управління в технічних системах та Навчально-наукового комплексу «Світовий центр даних з геоінформатики та сталого розвитку» Національного технічного університету

України «Київський політехнічний інститут імені Ігоря Сікорського». Матеріали дисертації були представлені, обговорювалися і одержали схвалення на міжнародних науково-практичних конференціях та форумах, зокрема: International Conference on Data Sharing and Integration for Global Sustainability (Нью-Делі, Індія, 2014); Міжнародна науково-технічна конференція «Системний аналіз та інформаційні технології» (Київ, 2014, 2015, 2018); 6th Digital Earth Summit “Digital Earth in the Era of Big Data” (Пекін, Китай, 2016); ISC World Data System Members' Forum (Денвер, США, 2016); International Data Week (Денвер, США, 2016); IEEE First Ukraine Conference on Electrical and Computer Engineering (Київ, 2017); IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (Бухарест, Румунія, 2017); IEEE International Conference on System Analysis & Intelligent Computing (Київ, 2018); 14th Federated Conference on Computer Science and Information Systems (Лейпціг, Німеччина, 2019); IEEE 2nd International Conference on System Analysis & Intelligent Computing (Київ, 2020); BSUN e-Conference on “Addressing the Impact of SARS CoV2 Pandemics on the Higher Education Activities in the Black Sea Region” (Константа, Румунія, 2020); ISC World Data System Members' Forum 2020 (Токіо, Японія, 2020).

Публікації. За темою дисертації опубліковано 17 наукових праць, у тому числі 5 монографій, 7 статей у наукових фахових виданнях (з них 1 стаття у періодичних наукових виданнях інших держав, які входять до ОЕСР та/або Європейського Союзу, 2 у фахових виданнях України категорії «А»), 5 публікацій у збірниках матеріалів конференцій.

Структура дисертації. Дисертація складається з переліку умовних позначень, вступу, чотирьох розділів, висновків, додатків і списку використаних джерел. Загальний обсяг дисертації становить 240 сторінок, а її основний зміст викладено на 195 сторінках. Робота містить 10 таблиць, 55 рисунків, 3 додатки і список використаних джерел із 154 найменувань.

РОЗДІЛ 1.

АНАЛІЗ ПРОБЛЕМ ОРГАНІЗАЦІЇ МІЖДИСЦИПЛІНАРНИХ ДОСЛІДЖЕНЬ В СВІТОВІЙ СИСТЕМІ ДАНИХ

Сьогодні ІТ стрімко розвиваються і, охоплюючи нові сфери діяльності людини, підтримують вирішення все складніших проблем, збільшують продуктивність праці. Важко уявити діяльність людини без використання ІТ. Фізичну працю людини давно виконують машини, якими керують комп'ютери, роботи. Але у реалізації функцій, пов'язаних з інтелектуальною діяльністю, комп'ютер виступає помічником. Крок за кроком він набуває все нових вмінь, прибираючи на себе все більше операцій, які раніше виконувала виключно людина. Однак у низці галузей, насамперед науковій діяльності, провідна роль належить людині. Тут раціонально використовувати ІТ можуть лише фахівці, які досконало знають свою предметну область і добре обізнані з їх можливостями [21]. Це ж стосується й міждисциплінарних досліджень в ССД.

У розділі виконано аналіз організації міждисциплінарних досліджень в ССД, визначена потреба у засобах їх підтримки, насамперед технологіях інтеграції джерел даних і застосунків, аналітичного оброблення даних, виявлені чинники впливу на вирішення зазначеної проблеми. Потім на основі аналітичного огляду існуючих технологій підтримки міждисциплінарних досліджень уточнена постановка проблеми створення, впровадження і ефективного використання платформи підтримки міждисциплінарних досліджень в ССД.

1.1. Міждисциплінарні дослідження: проблеми і перспективи

Розпочнемо зі Світової системи даних, її будови, завдань, ролі у розвитку науки. Система Світових центрів даних (Система СЦД, World Data Centers System) і Федерація астрономічних і геофізичних служб аналізу даних (ФАГС, Federation of Astronomical and Geophysical data analysis Services –

FAGS) створені Міжнародною радою з науки (MPN, International Council for Science – ICSU) в 1956 році в період підготовки до проведення найбільшого міжнародного наукового проекту XX століття «міжнародного геофізичного року» (1957-1958 рр.) для збору, зберігання, поширення і аналізу даних, одержуваних у багатьох областях науки.

Системи СЦД і ФАГС успішно служили міжнародній науковій спільноті понад 50 років [21]. За ці роки накопичено великі масиви даних і інформації, отримано знання і великий досвід з організації міжнародного обміну даними, аналізу та оброблення даних, створення архівів даних і забезпечення їх довгострокового зберігання, розвинено системи управління даними. Станом на 2008 рік функціонувало близько 50 Світових центрів даних і 13 Служб ФАГС, розташованих в 13 країнах на території Північної Америки, Європи, Азії та Австралії, у тому числі 15 у США, 8 у Японії, по 3 у Німеччині і Великобританії, 1 в Україні.

Однак ці системи на переломі тисячоліть виявилися не цілком адекватними потребам науки, недостатньо гнучкими для їх використання в міждисциплінарних дослідженнях. Тому в кінці 2008 р. на 29-й Генеральній асамблеї Міжнародної ради з науки в Мапуту (Мозамбік) було прийнято рішення інтегрувати існуючі структури Системи СЦД і ФАГС в нову міждисциплінарну структуру МРН – Світову систему даних (ССД, World Data System – WDS) для організації нового, скоординованого глобального підходу до наукових даних та інформації, який гарантує універсальний рівноправний доступ до якісних даних та інформації для досліджень, освіти та інформованого прийняття рішень.

Ще одним важливим кроком стало об'єднання у 2018 році Міжнародної ради з науки та Міжнародної ради з соціальних наук (International Social Science Council – ISSC, <http://www.worldsocialscience.org/>) в єдину організацію – Міжнародну наукову раду (International Science Council – ISC), що дозволило розширити спектр наукових дисциплін Світової системи даних за рахунок

залучення даних з області соціальних наук, включаючи економічні та поведінкові науки.

Роботу Світової системи даних координує Науковий комітет (World Data System Scientific Committee – WDS-SC). Найважливішими проблемами, які йому треба було розв'язати, були:

- уніфікація форматів і протоколів передачі даних;
- організація контролю якості наукових даних і інформації [22].

З метою узгодження дій і організації єдиного інформаційного простору для забезпечення збору, оброблення, обміну даними та вирішення фундаментальних і прикладних міждисциплінарних завдань до Світової системи даних увійшов СЦД з геоінформатики та сталого розвитку (СЦД-Україна), що працює на базі Інституту прикладного системного аналізу (ІПСА) Міністерства освіти і науки (МОН) України та Національної академії наук (НАН) України в структурі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (КПІ ім. Ігоря Сікорського, Київ, Україна) [23].

Розглянемо більш детально процеси створення і розвитку Світового центру даних в Україні. Україна була залучена до процесів збору глобальних даних ще за часів створення двох перших СЦД, коли значну частину роботи зі збору та підготовки даних виконували спеціалізовані академічні установи по всьому СРСР. З розпадом СРСР і появою нових країн протягом 90-х років виникла проблема ізоляції українських вчених від світового наукового співтовариства – колишні зв'язки з партнерами по обміну науковими даними зникали, а нові не встановлювалися.

Україна з початку свого існування змушена була вирішувати складне завдання – ліквідацію наслідків аварії на Чорнобильській АЕС. Рішення цілого спектру проблем привернуло увагу вчених з усього світу, що призвело до появи нових підходів в прогнозуванні та моделюванні природних і техногенних катастроф і сприяло відновленню мереж наукових спостережень і збору наукових даних.

Рішенням Президії НАН України, МОН України та Геофізичного центру РАН від 3 квітня 2006 року було створено Українське відділення Світових центрів даних з Сонячно-Земної фізики та фізики твердої Землі на базі ІПСА НАН України і МОН України в структурі КПП ім. Ігоря Сікорського.

Таке рішення визначило міждисциплінарну спрямованість українського СЦД, наявність власних синтезованих даних (результати досліджень з питань сталого розвитку, які проводить колектив ІПСА) і велика кількість наукових зв'язків ІПСА з науковими установами України та інших країн. Технічні і людські ресурси КПП ім. Ігоря Сікорського та ІПСА склали передумови для довгострокового забезпечення діяльності СЦД-Україна.

Ще одним важливим чинником розвитку СЦД в Україні стала ІТ-спрямованість КПП ім. Ігоря Сікорського, яка з самого початку дозволила на базі сучасних інформаційних технологій організувати процеси отримання, зберігання і передачі даних до СЦД-Україна.

Необхідно також зазначити, що КПП ім. Ігоря Сікорського є центральним вузлом Української науково-освітньої телекомунікаційної мережі URAN (Ukrainian Research & Academic Network), використання якої дозволило суттєво знизити витрати на передачу даних між науковими установами, що приєдналися до мережі [24]. Подальше підключення URAN до загальноєвропейської мережі GEANT2 в 2007 р відкрило ряд нових можливостей по обміну даними з науковими установами Європи і всього світу [24, 25]. Основні вузли мережі URAN і пропускна здатність каналів показані на рис. 1.1.

Значні обсяги даних і сучасні математичні моделі, які використовуються при дослідженні глобальних систем, обумовили потребу у новому якісному рівні обчислювальної техніки. Для цього в СЦД-Україна було створено обчислювальний центр, який забезпечує роботу кластера паралельних обчислень на базі архітектури Intel Xeon продуктивністю 7 ТФлопс і сховища даних, створеного за технологією IPStore. Складові загальної структури СЦД і структура їх взаємодії наведені на рис. 1.2.



Рис. 1.1. Українська науково-освітня телекомунікаційна мережа

У 2008 році Українське відділення Світових центрів даних з Сонячно-Земної фізики та фізики твердої Землі було сертифіковане як Світовий центр даних і приєднане до Системи Світових центрів даних в якості повноправного члена [22]. Згідно з отриманими рекомендаціями СЦД орієнтувався на формування глобального геоінформаційного простору для вивчення актуальних проблем про Землю. Відповідно вирішувалися проблеми його технічного забезпечення, закладалися перспективи довгострокової стабільності, відповідності нормам і зобов'язаннями Світових центрів даних та ін.

Перейдемо до опису моделі СЦД-Україна. Це мережна модель, структура якої відповідає рекомендаціям Світової системи даних. СЦД, як правило, мають яскраво виражену дисциплінарну спрямованість. Це пов'язано з тим, що однією з основних функцій, що регламентуються «Керівництвом для системи СЦД», є інформаційно-довідкова підтримка архівів і БД, створення та актуалізація інвентаризаційних каталогів, формування метаданих, надання консультацій користувачам [26], що є можливим лише за наявності в СЦД фахівців у відповідних наукових областях.

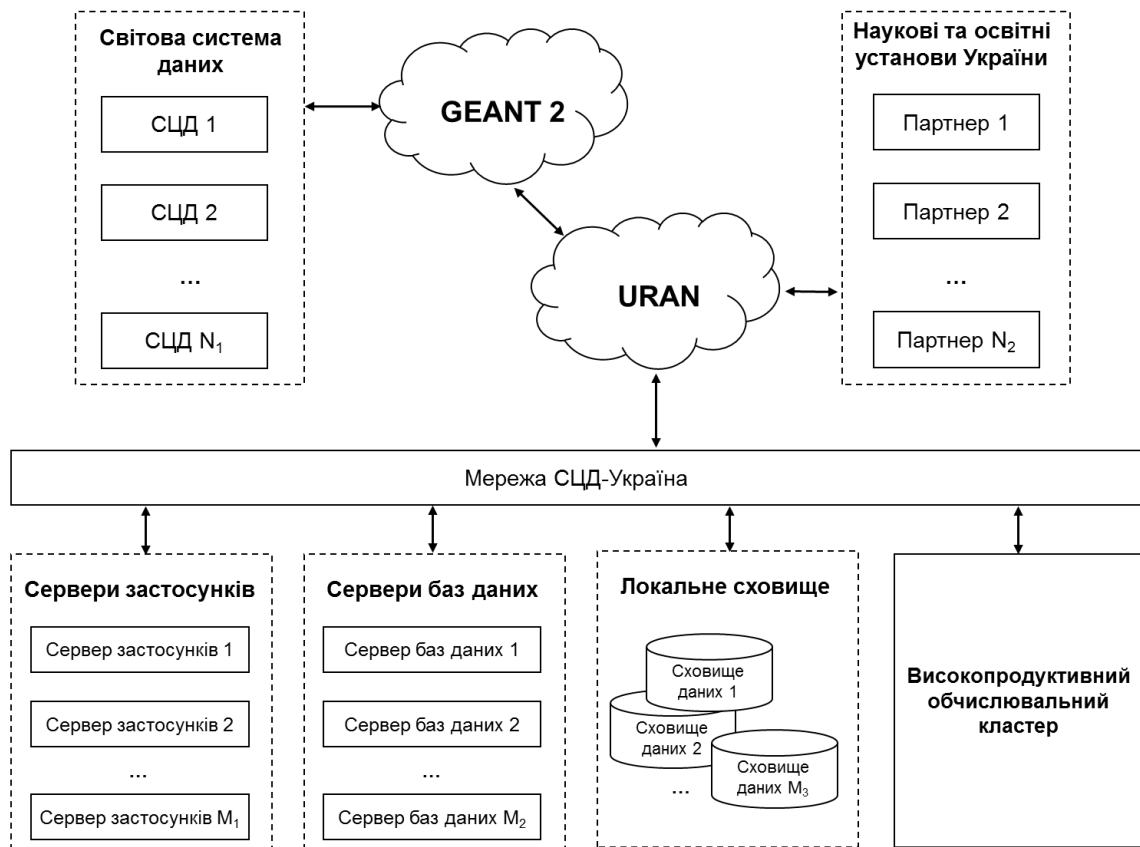


Рис. 1.2. Загальна структура взаємодії складових СЦД

СЦД в Україні покликаний забезпечити доступ українському науковому співтовариству до глобальних інформаційних ресурсів Міжнародної ради з науки в галузі наук про Землю, планетарної і космічної фізики і відповідних суміжних дисциплін, а також забезпечувати збір і зберігання національних наукових даних із зазначених дисциплін і їх репрезентацію світовому спільноті. На СЦД-Україна також покладено збір, оброблення та аналіз світових даних, необхідних для досліджень в галузі сталого розвитку [27, 28].

Для СЦД з геоінформатики та сталого розвитку, крім перелічених вище, були сформовані завдання шляхом встановлення договірних відносин з національними та світовими постачальниками даних:

- 1) організації ефективної системи збору інформації;
- 2) пошуку джерел даних;
- 3) розроблення підсистем прямого збору даних;
- 4) забезпечення зручного доступу для користувачів:
 - створення і підтримка веб-порталу з онлайн-доступом і пошуком даних (<http://wdc.org.ua>);

- забезпечення багатомовного інтерфейсу доступу до даних;
- проектування і розроблення баз даних;
- проектування і розроблення додаткового програмного забезпечення для зручної роботи з даними.

Врахування особливостей функціонування СЦД в інших країнах, міждисциплінарна спрямованість ІПСА та системний підхід дозволили запропонувати унікальну для світової системи даних мережеву модель функціонування СЦД-Україна як єдиного міждисциплінарного національного центру даних. Відповідно до цієї моделі кожна наукова дисципліна є сферою наукових інтересів однієї або кількох наукових організацій Національної академії наук України (рис. 1.3) [27]. Виділимо деякі з них:

- Інститут прикладного системного аналізу НАН України і МОН України (системне узгодження міждисциплінарних даних, сталий розвиток);
- Головна астрономічна обсерваторія НАН України (космічна геодезія та геодинаміка; космічні промені);



Рис. 1.3. Мережева модель СЦД-Україна

- Інститут геофізики НАН України ім. С.І. Субботіна (дані по сейсмології, гравіметрії, тепловому потоку, архео- і палеомагнетизму, магнітні вимірювання);
- Науковий центр аерокосмічних досліджень Землі Інституту геологічних наук НАН України (аерокосмічні знімки для використання в геології, екології, сільському, лісовому і водному господарстві, для прогнозування ризиків природних і техногенних процесів, прогнозування глобальних змін навколишнього середовища і катастрофічних процесів);
- Інститут географії НАН України (картографічні дані);
- Чорнобильський центр з проблем ядерної безпеки, радіоактивних відходів і радіоекології (дані щодо радіаційних, біологічних і медичних наслідків Чорнобильської катастрофи, безпеки «саркофагу»);
- Центральна геофізична обсерваторія (гідрологія, гідрографія, метеорологія та кліматологія, забруднення навколишнього середовища);
- ДУ «Інститут епідеміології та інфекційних хвороб ім. Л.В. Громашевського» НАМН України (дані з проблем інфекційної патології людини) та інші.

Мережеву модель було вперше представлено 7 жовтня 2009 року на спеціальній сесії СЦД "Emerging Technologies and Opportunities for Global Data Management and Exchange", де отримала схвалення і умовну назву «Network of networks» [29]. На засіданні Наукового комітету ССД 13-14 жовтня 2009 року в Парижі зазначена модель була прийнята як зразок для інших СЦД.

В Україні такий підхід, з одного боку, дозволяє ефективно використовувати існуючі технологічні можливості мережі URAN і високопродуктивний обчислювальний кластер КПІ ім. Ігоря Сікорського, а з іншого – сконцентрувати зусилля колективу СЦД на рішення міждисциплінарних завдань системного характеру, результати яких важливі

для всіх партнерів ССД. Сучасні дослідження складних систем передбачають їх одночасне вивчення з позицій багатьох наукових дисциплін, що становить передумову формування більш повної наукової картини світу на основі створення міждисциплінарних моделей, отриманих в результаті системного узгодження емпіричних даних, моделей, методів із різних наукових областей. Такі дослідження найчастіше ґрунтуються на взаємодії між багатьма їх учасниками на рівні засобів обміну і перетворення даних, технологій їх оброблення і аналізу, включаючи інструменти системного узгодження міждисциплінарних даних, їх систематизації, інтелектуального оброблення, оцінювання адекватності, аналізу якості, коректності тощо.

Саме ці задачі покладені на ССД. Однією з таких задач є глобальне моделювання процесів сталого розвитку в контексті якості та безпеки життя з метою отримання набору індексів та індикаторів сталого розвитку складних систем різних рівнів. Таке моделювання дає можливість моніторингу розвитку світу, окремих держав і регіонів, побудови сценаріїв розвитку і вироблення рекомендацій для осіб, котрі приймають рішення. Дослідження в цій галузі вимагають значної кількості різноманітних даних, велика частина яких відноситься до геоінформатики [12, 13]. З огляду на великий досвід і результати досліджень ІПСА (базової організації для ССД-Україна) в області сталого розвитку, Науковий комітет ССД доручив Світовому центру даних з геоінформатики та сталого розвитку підготовку щорічних звітів за результатами глобального моделювання процесів сталого розвитку в контексті якості та безпеки життя.

1.2 Сутність та особливості організації і підтримки міждисциплінарних досліджень в Світовій системі даних

Прогрес галузі ІТ спричинився до усталення і розвитку інформаційного середовища, переходу до розподілених обчислень, ефективного використання інформаційних та інших ресурсів і в галузі науки, насамперед для організації і підтримки міждисциплінарних досліджень в ССД. Консолідація ресурсів,

впровадження технологій віртуалізації призводять до заміни локальних рішень розподіленими, ефективного використання об'єднаних в глобальну мережу обчислювальних потужностей, систем збереження даних, забезпечуючи швидкий і надійний доступ науковців до накопичених інформаційних ресурсів.

Сервісний підхід, який сформувався на послугах зв'язку, поширився на інфраструктуру, платформи розроблення програмного забезпечення і власне програмні застосунки. Поява широкого спектру нових видів сервісів, насамперед контенту, призвела до конвергенції сервісів і формування узагальненого поняття інформаційно-комунікаційних сервісів (ІКС). З'явилися надійні провайдери інфраструктурних – Infrastructure as a Service (IaaS), платформних – Platform as a Service (PaaS) і програмних – Software as a Service (SaaS) сервісів. Функціональні можливості сервісів, вигідне співвідношення ціна/якість спонукають підприємства і організації відмовитися від розвитку власної ІТ-інфраструктури і перейти до компонентно-базованого проектування власних інформаційних систем з оптимальним вибором компонентів із широкого спектру наявних сервісів [30-32].

При цьому ефективне використання переваг розподілених систем виходить на новий якісний рівень, а сучасні сервіс-орієнтовані інформаційні технології забезпечують уніфікований доступ до сервісів [33]. Одночасно, формування нового демократичного ІТ-середовища дозволяє долучитися до розроблення і надання нових сервісів невеликим підприємствам і компаніям. Наукові установи не стоять осторонь. У такій ситуації, ефективно використовуючи різноманітні засоби і технології доступу, вони отримують нові можливості, насамперед потужні ресурси для організації і підтримки наукових досліджень. Міждисциплінарні дослідження в ССД отримують потужний імпульс [34].

ССД, яка забезпечує збір, збереження, поширення і аналіз даних, отриманих в багатьох галузях науки, перетворюється один із найпотужніших

інформаційних ресурсів [2]. У цій системі накопичено багато даних і застосунків, які можна використовувати для вирішення актуальних проблем розвитку суспільства. Але цьому перешкоджають труднощі, причини виникнення яких мають різну природу – політичну, соціальну, організаційну, фінансову, технічну, технологічну та ін.

Розглянемо сутність і особливості організації і підтримки міждисциплінарних досліджень в Світовій системі даних, щоб на їх основі виявити зазначені вище причини і побудувати платформу підтримки міждисциплінарних досліджень в Світовій системі даних, яка за рахунок драйверів розвитку інформаційних технологій нівелюватиме вплив зазначених причин і створюватиме умови розвитку міждисциплінарних досліджень.

В основі організації і підтримки міждисциплінарних досліджень в Світовій системі даних лежить забезпечення наукових досліджень у різних галузях науки і міждисциплінарних досліджень, спрямованих на розв'язання нагальних проблем розвитку суспільства.

Організаційна складова пов'язана з підтримкою процесів:

- планування власне наукової діяльності;
- координування спільних проектів;
- інформаційно-довідкової діяльності (пошуку і опрацювання потрібних дослідникам інформаційних ресурсів);
- збору, накопичення, оброблення і обміну потрібними інформаційними ресурсами;
- напрацюванням єдиної технічної і технологічної політики в галузі створення, впровадження, експлуатації і розвитку інформаційних систем різного рівня;
- підготовки наукових кадрів, здатних працювати в новому ІТ-середовищі.

Фінансова складова пов'язана з фінансуванням діяльності національних центрів, міжнаціональних структур і спільних проектів.

Технічна і технологічна складові пов'язані з реалізацією єдиної

технічної і технологічної політики в галузі створення, впровадження, експлуатації і розвитку інформаційних систем різного рівня.

Соціальна складова пов'язана з аналізом наукових проблем, розв'язання яких сприяє розвитку суспільства, оцінюванням внеску світової системи даних в розв'язання цих проблем, участю світової системи даних у формуванні ІТ-суспільства.

Політична складова пов'язана з аналізом стану міжнародних відносин і оцінкою та прийняттям рішень, які сприятимуть розвитку науки в поточних умовах.

Розглянемо більш детально зазначені складові, пов'язані з метою роботи.

Інформаційно-довідкова діяльність Світової системи даних (а це один із основних напрямів діяльності ССД), пов'язана з:

- 1) Підтримкою архівів і баз даних;
- 2) Створенням та актуалізацією інвентаризаційних каталогів;
- 3) Формуванням метаданих;
- 4) Наданням необхідних консультацій кінцевим користувачам.

Успадковані застосунки різних ССД мають різні можливості реалізації інформаційно-довідкової діяльності. Загалом користувач повинен бути готовий отримати потрібні дані як:

- великі файли (у цьому випадку користувач сам реалізує завдання з пошуку та оброблення даних);
- визначені користувачем поля записів бази даних, що відповідають умовам запиту (у цьому випадку користувач пише запити мовою відповідної СУБД, знаючи схему бази даних);
- визначені користувачем поля записів бази даних або файлів (у цьому випадку користувач заповнює зручні форми для вибору даних, які вимагають спеціальні застосунки-клієнти);
- визначені користувачем поля джерела (у цьому випадку користувач заповнює зручну форму вибору даних на основі метаданих джерела,

з якої автоматично генеруються запити).

Оскільки Світові центри даних, як правило, мають дисциплінарну спрямованість, зазначена діяльність можлива лише за наявності в СЦД фахівців у відповідних наукових галузях. Так, СЦД з геоінформатики та сталого розвитку (СЦД-Україна) в Україні покликаний забезпечити доступ українському науковому співтовариству до глобальних інформаційних ресурсів Міжнародної ради з науки в галузі наук про Землю, планетарної та космічної фізики та відповідних суміжних дисциплін, а також забезпечувати збір і зберігання національних наукових даних з зазначених дисциплін та їх репрезентацію світовому співтовариству.

На СЦД-Україна також покладено збір, оброблення та аналіз світових даних, необхідних для досліджень у сфері сталого розвитку.

У організацій-членів Світової системи даних накопичено багато даних і застосунків. Принциповою основою діяльності Світової системи даних є взаємодія СЦД, окремих наукових колективів і вчених. Вчені різних галузей знання в процесі міждисциплінарних досліджень використовують накопичені ресурси інших галузей. До кола джерел, доступ до ресурсів яких їм забезпечується, належать: бази даних; веб-сайти і портали; різноманітні успадковані файлові системи; репозиторії структурованих за різними моделями даних. Сьогодні ця взаємодія реалізується в рамках взаємодії успадкованих застосунків з усіма її особливостями, обумовленими відмінністю архітектур, різноманіттям форматів представлення даних та іншими чинниками.

Науковцям потрібен єдиний уніфікований інтерфейс доступу до джерел даних, який дозволить сприймати накопичені ресурси як нове єдине джерело. Для аналізу даних різноманітних джерел необхідна взаємодія успадкованих застосунків і швидка реалізація нового функціоналу. Постає потреба в проблемно-орієнтованих математичних моделях, методах і засобах отримання якісно нових даних, забезпечення взаємодії застосунків, отримання і узгодження даних з різних джерел.

У ССД є багато прикладів надання користувачам доступу до накопичених неоднорідних даних за допомогою кожного разу різної комбінації засобів та інструментів їх оброблення і аналізу [2]. Користувачі стикаються з труднощами при спробах отримати потрібні дані, якщо вони розподілені між декількома джерелами і потребують опрацювання із застосуванням декількох інструментів. У таких випадках користувач заздалегідь за допомогою ІТ-фахівців повинен отримати схему розв'язання своєї проблеми.

Настав час надати користувачам інструмент для комплексного розв'язання цієї проблеми, незалежно від типу і кількості джерел даних і успадкованих застосунків. Необхідно розробити ІТ, здатні забезпечити універсальний доступ до ресурсів. Накопичені ресурси, набутий досвід з організації міжнародного обміну даними мають скласти основу для рішень нового рівня, коли технології, формати, і процедури опрацювання даних не розділяють, а об'єднують наукову спільноту.

Розглянемо проблему підтримки міждисциплінарних досліджень більш прискіпливо. Щоб зробити це у наступному підрозділі, ми попередньо виділимо її основні підпроблеми і особливості, які необхідно врахувати при їх розв'язанні. По-перше, вимагають вирішення задачі інтеграції даних і застосунків. По-друге, постає задача оброблення запитів до розподілених джерел даних. По-третє, потрібно розробити комплекс моделей і методів системного узгодження даних і впровадити проблемно-орієнтований апарат інтелектуального оброблення даних. По-четверте, вимагає вирішення задача автоматизованого генерування проблемно-орієнтованих застосунків для аналітичного оброблення міждисциплінарних даних

1.3. Проблеми підтримки міждисциплінарних досліджень в Світовій системі даних

Отже, за понад 50 років Світові центри даних збирали, аналізували, архівували та поширювали дані широкого спектру спостережень. Центри забезпечують відкритий та зручний доступ до великих обсягів даних (метеорологія, океанографія, фізика, геоінформатика, сталий розвиток, ракети, супутники, обертання Землі тощо), постійно збільшуючи інформаційні ресурси в Інтернеті.

СЦД забезпечують довгострокове безпечне збереження та розповсюдження даних та продуктів на глобальному та регіональному рівнях. Наприклад, СЦД містять набори даних, що стосуються сонячної активності та міжпланетних явищ, іоносфери, геомагнітних змін, а також космічних променів, сейсмології, геомагнетизму і т.і. Зберігаються спостережувані дані за період від XVIII ст. Дані збираються з глобальних мереж спостережних станцій (дослідні судна, буї, супутники, спеціальні експерименти та інші пристрої) за допомогою телекомунікаційної системи ССД.

Відповідно виникають проблеми опрацювання зібраних даних. Дані моніторяться, перевіряються та оновлюються постійно. Дані доступні в різних традиційних формах, наприклад, на папері, мікрофільмах. Але сьогодні важливо зберігати дані в електронних форматах у зручній готовій для використання формі. СЦД перетворюють старі аналогові дані в цифрову форму.

Усі дані, що подаються до ССД, реєструються в комп'ютерних базах даних і заносяться в каталог доступності даних. Дані та інформація класифікуються та реєструються в каталогах та каталогах бази метаданих. Елементи метаданих дозволяють ідентифікувати тип та походження даних, їх просторове та часове покриття та інші характеристики, необхідні для забезпечення достовірності наборів даних. Метадані використовуються для підготовки електронних каталогів даних. Внутрішнє управління даними базується на стандартних а іноді оригінальних мовах описання даних (DDL) та

СУБД. Цифрові дані у нестандартних форматах, метаданих та каталогах доступності даних доступні у вільному доступі на веб-сайтах СЦД.

Важливою проблемою СЦД є аналітичне оброблення даних з метою надання фахівцям додаткових продуктів аналізу даних, наприклад карт розподілу даних, прогнозів, аналітичних оцінок. Спеціалізовані високоякісні набори даних для вивчення клімату мають особливе значення та надаються для доступу в Інтернет.

У ССД застосовуються процедури оцінювання та контролю якості даних, контролю регламентів обслуговування даних. Для перевірки використовуються набори візуальних та автоматичних процедур контролю якості відповідно до керівництв WMO (World Meteorological Organization) і ІОС (Intergovernmental Oceanographic Commission). Кількісні критерії, які використовуються в процедурах контролю якості, оновлюються на основі постійних досліджень. Використовуються процедури перевірки дублікатів для уникнення дублікації даних у глобальних наборах даних. Для забезпечення цілісності метаданих каталоги організацій, дослідницьких проектів регулярно перевіряються на відповідність міжнародним каталогам колег IODE (International Oceanographic Data and Information), GCMD (Global Change Master Directory), ICES (International Council for the Exploration of the Sea), WMO (World Meteorological Organization), EDMO (European Digital Media Observatory) та EDMERP (European Directory of Marine Environmental Research Projects). Усі ці заходи забезпечують певний рівень ідентичності даних та метаданих, цілісність та ефективне архівування та використання даних.

Для досягнення високої швидкості оброблення даних та забезпечення ефективного доступу користувачів до даних в Інтернеті застосовується комбінована стратегія впровадження сучасних інтегрованих web-технологій.

Для доступу до накопичених даних та результатів їх аналітичного оброблення використовуються поширені СУБД, насамперед Oracle. Це дає можливість користувачеві знайти потрібні дані, отримати їх, переглядати та представляти у вигляді таблиці, сюжету, карти чи діаграми. Формування та

розповсюдження міжнародних видань (атласів, газет, довідників, посібників та вказівок) також входить в сферу відповідальності СЦД. Каталоги міжнародних видань доступні в Інтернеті.

СЦД суворо дотримуються міжнародних стандартів описання метаданих, обміну даними. Для розміщення web-метаданих, що відповідають міжнародним профілям метаданих, використовується стандарт ISO 19115: основний профіль WMO для даних метеорології [35] та CDI для даних океанографії [36]. Для обміну даними використовуються формати даних BUFR/CREX (від WMO), та поширені формати даних NetCDF та ODV.

Кожний СЦД використовує ІТ-інфраструктуру, яка відповідає вимогам до процесів збору, довгострокового збереження, оброблення та розповсюдження даних. Хоч історично технічні засоби (комп'ютери, сховища, комунікаційні засоби) і програмне забезпечення відрізняються від одного СЦД до іншого, стандартні процедури (архівація, контроль умов, періодичні перевірки, резервне копіювання, відновлення, дзеркалювання, шифрування та інші), застосовуються як до глобальних даних та інформації, так і до національних даних та інформації. Для оброблення даних історично використовуються мейнфрейми IBM z9 BC, а також різні суперкомп'ютерні рішення, для збереження даних – сховища IBM System Storage DS8300, IBM System Storage 3500. Висока продуктивність, велика ємність (до 512 Тб), система брандмауера, повне шифрування та інші процедури забезпечують високий рівень безпеки даних.

Однак СЦД постають перед низкою проблем, які вирішені частково, або ще не вирішені. Покажемо ці проблеми на прикладі українського СЦД, який має доступ до інформаційних ресурсів МНР у багатьох галузях науки, одночасно збираючи, накопичуючи, оброблюючи і поширюючи дані української наукової спільноти в тих же галузях. СЦД-Україна також збирає, оброблює, аналізує і поширює глобальні і національні дані, пов'язані з дослідженнями сталого розвитку.

Здійснюючи фундаментальні і прикладні дослідження, аналіз і

вирішення міждисциплінарних проблем системного характеру, насамперед кількісне вимірювання і моделювання процесів сталого розвитку, оцінювання впливу загроз на розвиток в глобальному і регіональному контекстах, ССД-Україна також вирішує низку проблем, пов'язаних з підтримкою міждисциплінарних досліджень. Насамперед мова йде про проектування, реалізацію і впровадження інформаційних технологій, що створюють умови безперешкодного збору, оброблення, обміну, аналізу міждисциплінарних даних з метою забезпечення умов для розвитку. Щоб зберегти позицію одного із лідерів у дослідженнях сталого розвитку, український ССД з питань геоінформатики та сталого розвитку визначив такі пріоритети:

- 1) інтеграція з новими ССД та ефективне співробітництво з членами Світової системи даних;
- 2) забезпечення безпеки даних на нецифрових носіях інформації та їх оцифрування;
- 3) прийняття і дотримання політики забезпечення якості даних;
- 4) розвиток інфраструктури провайдерів даних;
- 5) формування загального інформаційного простору, забезпечуючи інтеграцію джерел даних і застосунків;
- 6) створення інструментарію аналітичного оброблення інформації.

Одним з найважливіших завдань у рамках визначених пріоритетів визнане створення загального інформаційного простору для Світових центрів даних МНР. І мова йде не лише про представлення даних в уніфікованих форматах. Наступним кроком є розроблення уніфікованих інструментів для ефективного обміну даними. З одного боку, такий підхід дозволить створити єдину точку доступу до всіх ресурсів, даних, сервісів та служб ССД. З іншого боку, це також забезпечить гнучку основу для розроблення єдиного інструментарію оброблення даних. Основними рисами створюваного загального інформаційного простору вбачаються:

- 1) Гнучка та масштабована міжплатформна архітектура з відкритим кодом, наприклад, SOA;

- 2) Наявність централізованого реєстру даних та послуг;
- 3) Проста інтеграція з існуючими системами, наприклад за допомогою SOAP, WSDL, BPEL, UDDI тощо;
- 4) Наявність єдиної точки доступу;
- 5) Наявність легко створюваного та налаштовуваного інтерфейсу користувача, заснованого на існуючих послугах;
- 6) Наявність універсального спільного підходу до отримання даних з різних джерел даних.

Усі СЦД виявили зацікавленість у приєднанні до нової Світової системи даних МНР та успішно пройшли необхідні етапи сертифікації.

В Україні для реалізації такого підходу є технологічні можливості згадуваної вище мережі URAN, яка об'єднує наукові і освітні заклади, і потужний суперкомп'ютер КПП ім. Ігоря Сікорського, які формують основу для співпраці, розподілу завдань між партнерами. Крім того, систематичний акцент на міждисциплінарних дослідженнях організацій і колективів СЦД-Україна дозволив накопичити досвід, який без сумніву буде корисний для усіх членів ССД:

- забезпечення всіх основних етапів управління науковими даними (збір, забезпечення якості, зберігання, оброблення, обмін, звітність та довгострокове управління) різного виду із різних сфер і галузей науки;
- розроблення математичних моделей, методів та інструментів оцінювання та прийняття рішень у складних системах;
- розроблення та підтримка інформаційних систем та послуг, орієнтованих на аналіз та оброблення даних.

Підтримка міждисциплінарних досліджень у Світовій системі даних виглядає як велика науково-практична проблема, яка вимагає ефективного вирішення. По суті вона декомпозується на декілька підпроблем різного рівня складності, насамперед інтеграцію застосунків і джерел даних, раціональну організацію збереження, оброблення і поширення даних, створення моделей

узгодження даних, впровадження проблемно-орієнтованого апарату інтелектуального аналізу даних та формування ІТ-спільноти, здатної виконувати складні міждисциплінарні дослідження.

Стисло розглянемо зазначені підпроблеми, розуміючи, що мова все ж повинна йти про загальне рішення, яке інтегрує існуючі розв'язання підпроблем в єдиній платформі підтримки міждисциплінарних досліджень у Світовій системі даних.

По-перше, у ССД вже давно постала перед багатьма організаціями *підпроблема інтеграції даних і застосунків*. Якщо дотепер технологічні можливості в цій сфері були досить обмеженими, то сьогодні склалися умови для її ефективного вирішення. Стрімкий розвиток ІТ на переломі тисячоліть змінив традиційні уявлення про інформаційні процеси. Хмарні обчислення, сервісний підхід створили умови для консолідації програмно-апаратних засобів, конвергентності сервісів, віртуалізації ресурсів. Понад 50 функціонуючих на сьогодні ССД не повною мірою інтегрувалися в нове середовище. Їх потужні системи накопичення і оброблення інформації виявилися не цілком адекватними потребам науки, недостатньо гнучкими для їх використання в міждисциплінарних дослідженнях.

Тому перед ССД постало завдання впровадити новий, координований глобальний підхід до організації збереження, оброблення і обміну науковими даними та інформацією, який гарантує універсальний рівноправний доступ до якісних даних та інформації для досліджень, освіти та прийняття рішень. При розробленні моделей, методів і технологій інтеграції даних і застосунків треба врахувати її особливості, насамперед історичні дані, великі масиви даних, різноманіття форматів, дані переважно є результатом вимірювань. Потрібно забезпечити універсальний доступ до різноманітних джерел даних і взаємодію успадкованих застосунків, для яких характерними особливостями є відмінність архітектур, різноманіття форматів представлення даних та інші негативні в умовах формування нового ІТ-середовища чинники.

Використовуючи переваги новітніх ІТ необхідно насамперед

уніфікувати формати і протоколи передачі даних, організувати контроль якості наукових даних та інформації [2].

Зусилля повинні спрямовуватися на те, щоб вчені різних галузей знання могли використовувати накопичені ресурси, орієнтуючись на знання предметної області, а не особливості ІТ. До кола джерел, інтегрований доступ до ресурсів яких доцільно забезпечити, належать бази даних, веб-сайти і портали, різноманітні успадковані файлові системи і репозиторії структурованих за різними моделями даних.

Зазвичай інтеграцію даних в інформаційних системах розуміють як забезпечення єдиного уніфікованого інтерфейсу для доступу до певної сукупності, у загальному випадку, неоднорідних незалежних джерел даних [37]. Для користувача інформаційні ресурси всієї сукупності джерел, які інтегруються, повинні представлятися як нове єдине джерело, а система, яка забезпечує такі можливості, називається системою інтеграції даних [2].

Погляд на інтеграцію як взаємодію застосунків не такий поширений, але дуже важливий з точки зору вирішення комплексної проблеми інтеграції в новому ІТ-середовищі. Загалом він формувався у середовищі фахівців з автоматизації програмування [37], але поступово через проблематику інтеграції мультибаз і федеративних баз даних, великих сховищ даних, різноманітних репозиторіїв був перенесений на інтеграцію веб-застосунків [2]. Цей напрям сьогодні видається надзвичайно важливим, оскільки в умовах сервісного підходу його можна розглядати як теоретико-методологічну основу впровадження компонентно-базованого підходу до створення інформаційних систем.

Тому інтеграція даних в інформаційних системах і інтеграція як взаємодія застосунків мають комплексуватися. Виникає необхідність у розробленні технічного рішення, яке дозволить:

- створювати механізми ефективного управління інформаційними ресурсами;

- розширювати сфери стандартизації протоколів і форматів обміну даними на основі взаємодії відкритих систем;
- створювати крос-платформні застосунки, інтегрувати дані і застосунки;
- спрощувати розгортання і розвиток систем та ресурсів за рахунок приховування внутрішніх деталей апаратного і програмного забезпечення.

Необхідно створити математичні моделі і методи для технічної і технологічної реалізації інтеграції застосунків і джерел.

В умовах згаданої вище дисциплінарної спрямованості Світових центрів даних впровадження сервісного підходу в такій великій розподіленій системі має особливості, які треба врахувати при створенні платформи. Етапи еволюції програмних сервісів надання даних у СЦД [26]:

1) дані надавалися користувачу просто як великі файли, причому усі завдання з пошуку та оброблення даних у ньому користувач повинен був виконувати самостійно;

2) збережені СЦД дані було перенесено в бази даних, а користувачеві було надано засоби для введення запиту на них. Такий підхід вимагав від користувача знання мови запитів відповідної СУБД та інформації про схему БД, до яких необхідно виконати запит;

3) клієнт-серверні технології локальних мереж обумовили в СЦД появу «тонких» та «товстих» застосунків-клієнтів, які надавали користувачам форми для вибірки даних. Вони інкапсулювали в собі засоби генерування запитів до джерела даних, забезпечували користувачеві дружній інтерфейс, але розроблення таких форм часто вимагало істотних трудовитрат і, в більшості випадків, – індивідуального підходу до кожного джерела (або файлу) даних;

4) форми вибору даних еволюціонували у фасетні форми пошуку даних, у яких форми пошуку даних автоматично генеруються для кожного джерела даних і можна використовувати додаткові описи для збереження метаданих кожного джерела.

5) створення системи сервісів надання даних за принципом «єдиного вікна» для доступу до даних. Така організація передбачає інтеграцію джерел даних в єдину систему з урахуванням їх семантики. Важливі на етапах 1 – 4 каталоги посилань до засобів надання даних та їх метаописи, необхідні для структурування доступних джерел даних, з яких розпочинав взаємодію із сервісом надання даних користувач, перетворюються у невід’ємну частину власне сервісу надання даних. Втілення цієї ідеї породжує наукову проблему інтеграції інформаційних ресурсів різнорідних джерел.

Підпроблема оброблення запитів до розподілених джерел даних природно декомпонується на фундаментальні задачі побудови формалізмів для метаописання джерел і формування запитів, їх розбору (парсингу), формування і оптимізації плану виконання запитів [26]. Вони виникають у будь-яких системах зберігання даних – розподілених, централізованих чи паралельних – оскільки оброблювач запитів має отримати запит, зрозуміти значення його параметрів, сформулювати план виконання запиту, оптимізувати і виконати його, щоб надати потрібні користувачам і застосункам дані.

Формальні засоби метаопису джерел і формування запитів звичайно визначаються на основі логічних, алгебраїчних і графічних формалізмів. Метаопис має зберігати необхідну для формулювання, парсингу, перетворення і оптимізації запиту інформацію. Для цього він має містити схему джерела (джерел) даних (наприклад, визначення таблиць, представлень, типи даних і т.і.), схеми поділу (наприклад, інформацію про те, що глобальні таблиці були розділені і як вони можуть бути відновлені), фізичну інформацію (розташування розділів таблиць, структуру індексів і статистику для оцінювання ефективності виконання плану).

Поданий у зручному для користувача вигляді запит має бути обробленим і переведеним у внутрішнє представлення, наприклад, граф запиту, зручне для парсингу.

Формування плану виконання запиту, насамперед, передбачає його еквівалентне перетворення, спрямоване на ліквідацію надлишкових

предикатів, спрощення виразів і усунення вкладених запитів. На виході отримуємо план, який визначає послідовність виконання підзапитів до джерел даних та стратегію їх об'єднання. А для розподілених систем визначається, які частини запиту будуть виконуватися на яких джерелах даних.

Оптимізація плану обумовлена необхідністю врахувати при його виконанні фізичний стан системи – розмір таблиць, індекси, швидкість процесора тощо. Вона передбачає уточнення індексів і методів (наприклад, хешування і сортування), які будуть використовуватися для виконання запиту і його операцій (наприклад, об'єднання і групування), і в якому порядку вони будуть застосовані. Також уточнюються вузли, на яких виконуються операції, остаточно формується послідовність виконання підзапитів до джерел даних і стратегія їх об'єднання. Звичайно плани подаються у вигляді дерев, вершинами яких є оператори (з'єднання, групування, сортування та ін.), а ребра вказують послідовність виконання операторів.

Оптимізація запитів у гетерогенній системі має бути універсальною і базуватися на оцінках «вартості» виконання запитів до джерел даних. Застосовуються кілька стратегій оптимізації [38]. Найчастіше будується узагальнена модель оцінювання вартості виконання запиту для всіх компонентів і її коефіцієнти корегуються для кожного компоненту за результатами виконання набору тестових запитів. Рідше будується своя модель оцінювання вартості для кожного компоненту. Такий підхід дає найвищу точність оцінок вартості виконання запитів, але його реалізація є найвитратнішою. Практично доцільним для вже функціонуючих розподілених систем є визначення вартості плану виконання запитів на основі статистичних даних моніторингу виконання компонентами запитів.

По оптимізації постає задача трансформації плану виконання запиту у виконуваний план. У задачах міждисциплінарних досліджень ефективно виконання окремих частин плану, наприклад предикатів і виразів, на джерелі даних може включати специфічні розв'язання, такі як генерація асемблероподібного коду.

Задача власне виконання запиту у ССД полягає у використанні загальних реалізацій для кожного оператора. Особливості міждисциплінарних досліджень схиляють базувати організацію роботи підсистем виконання запитів на моделі ітератора, яка забезпечує реалізацію операторів і стандартний інтерфейс.

Загалом, виходячи із необхідності збереження дисциплінарності ССД, для інтеграції їх різнорідних джерел даних найперспективнішим видається підхід, реалізація якого дозволить звертатися до даних, трансформуючи їх у сумісні формати у реальному часі і не зберігаючи в проміжних сховищах. Це дозволить доповнити діяльність ССД широкими можливостями обміну даними і доступу кінцевих користувачів до ресурсів усіх ССД, незалежно від їх розташування, форматів даних і програмних засобів аналізу і оброблення даних.

Підпроблема системного узгодження даних і впровадження проблемно-орієнтованого апарату інтелектуального оброблення даних. Насамперед мова йде про узгодження даних на рівнях семантики, єдиного діапазону значень і єдиних одиниць вимірювання за умови мінімізації інформаційних втрат такого узгодження. Апарат щонайменше має включати моделі і методи сталого розвитку, класифікації і виявлення подібності.

Специфіка ССД стосовно *підпроблеми автоматизованого генерування проблемно-орієнтованих застосунків для аналітичного оброблення міждисциплінарних даних* виявляється переважно у потребі формалізації узагальнення різнорідних даних, які описують окремі аспекти дуже складних явищ і систем, використанні універсальних моделей прийняття рішень, оцінювання і візуалізації оброблюваних даних і результатів.

Тепер можемо переходити до огляду існуючих рішень.

1.4. Огляд існуючих рішень

Розглянемо рішення, які можна використати для підтримки міждисциплінарних досліджень СЦД ССД. Розпочнемо з інтеграції застосунків і джерел даних. Термінологічну основу інтеграції закладено у відомому звіті ANSI/X3/SPARC [39]. Різноманіття існуючих систем, підходів, технологій і засобів інтеграції систематизуватимемо за такими ознаками:

а) вид інтеграції джерел даних: ЕІІ (Enterprise Information Integration); ETL (Extract, Transform, Load); ЕАІ (Enterprise Application Integration);

б) спосіб організації взаємодії джерел даних: з посередником; peer-to-peer (P2P);

в) місце розташування описів джерел даних (каталогу): на центральному вузлі; повної копії загального каталогу на кожному вузлі; зберігання на кожному вузлі каталогу тільки тих об'єктів, які на ньому зберігаються; гібридний варіант;

г) архітектура програмної реалізації інтегрованої інформаційної системи: традиційна; клієнт-серверна; web-орієнтована;

д) рівень описання і маніпулювання даними: синтаксичний; семантичний;

е) метамодель описання інтегрованих даних: базова (реляційна, об'єктно-орієнтована, об'єктна); моделі класу NoSQL; простір даних; моделі дескриптивної логіки (в термінах OWL, RDF); структуровані моделі на основі XML; гібридні (SQL/XML) та ін.

Для оцінювання систем інтеграції інформаційних ресурсів виберемо традиційні критерії [40]: гетерогенність; розподіленість; автономність. За додаткові критерії їх оцінювання приймемо доступність, цілісність і ефективність. Розглянемо особливості систем інтеграції за наведеними ознаками.

Розпочнемо з класів систем за *видом*. Система ЕІІ підтримує інтеграцію даних в режимі реального часу на основі загального шлюзу (gateway) з єдиними мовою і точкою доступу до неузгоджених джерел даних (за рахунок

віртуалізації застосункам надається гнучкий, незапланований, з будь-якою періодичністю і обсягами доступ до даних). Система ETL підтримує пакетну інтеграцію, забезпечує вилучення даних із гетерогенних джерел, їх трансформацію і завантаження в єдине сховище даних (ефективна в сховищах добре документованих надійних даних для багатовимірного аналізу, наприклад, часових рядів, забезпечуючи видалення дублювання, перевірку якості даних, обслуговування співробітників і підтримку бізнес-процесів). Система EAI підтримує інтеграцію програм, координована взаємодія яких забезпечує перетворення даних (забезпечуючи взаємодію застосунків в реальному часі для автоматизації бізнес-процесів та узгодженість при внесенні змін у застосунки) [41, 42].

Важливим для розуміння сутності інтеграції джерел даних є їх поділ за *способом організації взаємодії джерел даних*. В системах інтеграції з посередником клієнти взаємодіють з системою через компонент-медіатор як показано на рис. 1.4 [38]. Медіатор здійснює розбір, переписування та оптимізацію запиту, виконує деякі з його операторів [43]. На основі каталогу, у якому зберігається глобальна схема гетерогенної системи баз даних, встановлюється відповідність її частин кожному компоненту бази даних, формулюються запити прикладних програм і користувачів, зовнішніх компонентів, оптимізуються запити.

Щоб об'єднати деталі компонентів БД, створюються обгортки («wrapper») для кожного компоненту, що не є складною проблемою. Обгортки транслюють кожен запит медіатора, щоб він був зрозумілий цільовій БД, і одержувані від БД результати, щоб вони були зрозумілі медіатору і сумісні з зовнішньою схемою БД і глобальною схемою гетерогенних баз даних. Розширюваність забезпечується тим, що обгортки та компоненти БД можуть бути оновлені або нові компоненти БД можуть бути інтегровані без зміни посередника та існуючих обгортки.

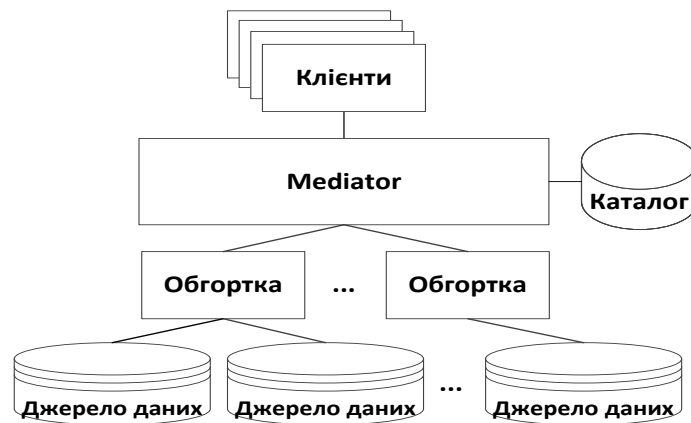


Рис. 1.4. Загальна схема системи інтеграції з посередником

Реалізації на основі медіаторів, насамперед, DISCO, Garlic, IBM Data Joiner, інструментів для побудови розподілених систем баз даних продемонстрували їх високу ефективність, особливо для інтеграції реляційних і об'єктно-орієнтованих баз даних. Вдало використовуючи переваги трирівневих архітектур програмних систем, вони ефективно інтегрують розподілені семантизовані джерела даних [38, 44-47].

Структуру систем інтеграції даних на основі взаємодії P2P [48] складають автономні вузли (джерела даних), пов'язані між собою за рахунок відображень. Вузол надає свою частину інформації, доступної з розподіленого середовища. Ці системи є гнучкими, децентралізованими і динамічними, оскільки вузли можуть підключатися або відключатися, не перериваючи роботу всієї системи.

У цих системах кожний вузол є системою інтеграції даних з медіатором, управляючи локально пов'язаними джерелами даних (локальний мапінг) за рахунок трансформованих в глобальну вузлову схему локальних відображень (глобальний мапінг). Вузли запрошують інформацію у будь-якого вузла (від зовнішніх користувачів або інших вузлів) на основі відображень, які зберігаються в їх специфікаціях. Опитуваний вузол завдяки мапінгу для формування відповіді може використовувати дані інших вузлів. Покладаючись тільки на доступні у вузлах сервіси, вузли системи ефективно надають відповіді на запити в термінах їх схем даних. Конвертація цих відповідей в єдиний формат на основі відображень між вузлами є складною,

як і обмін даними і відображеннями даних між вузлами.

За ознакою *місця розташування описів джерел даних виділяють* системи з централізованим описом джерел даних (метаопис, каталог для збереження метаданих), який дозволяє одним вузлам звертатися до об'єктів БД, що містяться на інших вузлах. Такі рішення незадовільні за критерієм автономності, оскільки порушують принцип незалежності вузлів від центрального вузла або частково усіх вузлів, адже будь-яке оновлення каталогу повинно виконатися на всіх вузлах системи. Локальний варіант зберігання незадовільний за критерієм ефективності, оскільки пошук об'єкта вимагатиме звернення в середньому до половини вузлів.

Поєднання цих підходів може бути ефективнішим. Так, на кожному вузлі промислової федеративної СУБД IBM System R зберігалася інформація про локальні об'єкти даного вузла і частина інформації загального каталогу – інформація про створені на цьому вузлі об'єкти. Складові системного імені кожного об'єкта однозначно ідентифікували його в федеративній базі даних [49].

Для досягнення критеріїв інтеграції класифікація за архітектурою програмної реалізації інтегрованої інформаційної системи є визначальною. Успадковані архітектури – від файлових систем до пакетів програм – мають слабкі інтеграційні можливості і їх включення до множини джерел вимагає багатьох зусиль. Клієнт-серверні трирівневі архітектури зручні для інтеграції на синтаксичному рівні баз даних, особливо реляційних і об'єктно-реляційних і меншою мірою об'єктних. Web-орієнтовані архітектури зручні для інтеграції на семантичному рівні, оскільки мають стандартизовані засоби обміну даними і взаємодії в термінах онтологій.

Переваги і недоліки систем інтеграції з *синтаксичним і семантичним рівнями* описання і маніпулювати даними визначаються їх орієнтацією на однорідні набори даних (синтаксичний рівень) і зв'язки та відношення між одиницями даних без залежності від їх подібності (семантичний рівень) [50]. У перших на запит одних джерел вибираються і перетворюються у

відповідності із заданими вимогами дані з інших джерел, а вартість з кожним новим джерелом даних зростає експоненційно. У других встановлюються зв'язки між одиницями даних у відповідності до визначень у їх онтологічних описах, а вартість практично не залежить від кількості джерел, оскільки у цьому випадку потрібно лише підготувати описи нових джерел. До того ж, тут зникає потреба у завантаженні даних у проміжні сховища. Жорсткі стандарти, формати синтаксичної інтеграції, порушення яких призводить до втрати контексту, поступаються гнучкому дотриманню стандартів, можливості введення унікальних типів у межах кожного опису джерела даних при семантичній інтеграції.

Насамкінець розглянемо особливості систем інтеграції, виділених за ознакою *метамоделі описання інтегрованих даних*. Оскільки більшість з них широко відомі, стисло охарактеризуємо концепції семантичної інтеграції даних, побудовані на методах визначення семантики предметної області. Інтеграція гетерогенних інформаційних ресурсів вимагає їх інтероперабельність на семантичному рівні, що породжує задачі порівняння вмісту ресурсів, відшукування відповідностей і розв'язування конфліктів, сполучення неоднорідних ресурсів [51].

Повна специфікація ресурсу повинна охоплювати специфікації: власне джерела даних; обмежень цілісності; контексту (області використання ресурсу).

Структура і поведінка інформаційного ресурсу визначена семантикою його предметної області. Умовою ефективного використання семантики даних є подання їх специфікації формальною мовою. Теоретичну основу такої мови складає мова числення предикатів першого порядку. Розвиток web-технологій сприяв появі мов представлення знань на основі дескриптивної логіки [52].

Як інструмент створення таких узагальнених специфікацій можна використати онтологічні специфікації, розуміючи під онтологією формалізований опис загальноприйнятого розуміння деякої предметної області, за допомогою якого можуть спілкуватися люди, комп'ютерні системи

[53]. Для інформаційного компонента така специфікація становить набір визначень і понять, а також правил (аксіом), пов'язаних з визначеннями і поняттями предметної області.

Онтологія визначає загальний словник для спілкування з використанням інформації предметної області. На відміну від метаданих (тип, розмір атрибута), онтологія має виразні засоби визначення семантики даних – ієрархію понять і типів об'єктів, описи типів, аксіоми, які задають обмеження на інтерпретації цих понять.

Розроблення і інтеграція програмних систем базуються на стандартних онтологіях, що дозволяє використовувати інформацію предметних областей людьми і комп'ютерами. Існуючі онтології уможливають повторне використання, аналіз і класифікацію знань, явне формулювання припущень в предметній області.

Через онтології визначаються набори даних та їх структури для використання іншими програмами. Тому для опису онтологій необхідна формальна мова. Обмін програм і програмних агентів даними, описаними в термінах онтологій, забезпечує їх реальну взаємодію, незалежність від форматів представлення даних. Потреба в онтологічних моделях на основі логік стимулюється реалізацією Semantic Web – концепції розвитку Інтернет, підтримуваної World Wide Web Consortium (W3C). Розроблено багато мов описання онтологій, наприклад Knowledge Interchange Format, Common Logic, Cycl, Simple HTML Ontology Extension (SHOE), Ontology Markup Language (OML). Подання зрозумілого для комп'ютерів семантичного знання дозволяє програмам і програмним агентам збирати значиму інформацію на web-сторінках і документах, полегшує пошук і опрацювання знань. Ядерна частина цих мов пов'язана з логічними аспектами, а онтологія визначається як структура, компоненти якої охоплюють декілька рівнів представлень, наприклад концептів, об'єктів, відношень, функцій, аксіом і онтологій, використовуючи концептуальні графічні можливості. Найбільш поширені такі основні мови:

1) XML (eXtensible Markup Language) – метамова для web-орієнтованих програмних систем і інтероперабельності з Standard General Markup Language (SGML) та HTML. Зручний для програм і людей синтаксис дозволяє визначати складні структури даних, опрацьовувати документи, перетворюючи її в мову специфікацій онтологій і засіб обміну онтологіями [54];

2) DARPA Agent Markup Language + Ontology Interchange Language (DAML+OIL) – сформований на стандартах W3C результат злиття мов DAML і OIL, розроблений для надання семантичної інтероперабельності в XML. Розширений набір потужних примітивів моделювання використовується для формування онтологій з формальною семантикою і сервісами логічного виведення на основі дескриптивної логіки, обміну онтологіями.

3) Ontology Web Language (OWL) [55] – мова описання Web-онтологій, розроблена як розширення Resource Description Framework (RDF) і RDF Schema (RDFS). OWL забезпечує поряд з формальною семантикою додатковий словник, що закладено в основу DAML + OIL, а також вбудовану підтримку відображення онтології. Її діалекти (Lite, DL і Full) відрізняються формальною складністю і виражальними можливостями – від класифікаційної ієрархії і простих обмежень до максимальної виразності і застосування аксіом і правил для логічного виведення і складними обмеженнями та синтаксичною свободою RDF без гарантій обчислення. Вони забезпечують потреби створення програмних систем інтеграції і підтримки взаємодії застосунків і через них – бізнес-процесів у семантичних термінах [56].

Огляд мов описання онтологій для Semantic Web дозволяє зробити висновок, що вони відповідають потребам створення програмних систем інтеграції.

Розглянемо тепер засоби семантичного описання інформаційних ресурсів та виконання запитів до них. До перших належить рекомендована W3C мова RDF – абстрактна модель для описання ресурсів у вигляді орієнтованого позначеного графа, причому кожний ресурс може мати властивості, які, в свою чергу, також можуть бути ресурсами або їх колекціями

[57, 58]. RDF, забезпечуючи універсальний, гнучкий метод декомпозиції знань на триплети і враховуючи їх семантику, дозволяє задати структуру опису джерела (бази даних, XML-документу, Web-сайту тощо). RDF використовується як для описання структури (вбудовані поняття RDF-схем – класи, властивості, типи, колекції, успадкування класів і властивостей), так і для визначення і оперування твердженнями на основі предикатів, що дозволяє подавати концептуальну інформацію.

Найпопулярнішою серед других є мова запитів до RDF-сховищ SPARQL Protocol and RDF Query Language [59]. Вона визначає тільки запити до інформації, що міститься в моделях, а реалізацію інтерфейсу для мови запитів забезпечує бібліотека для розроблення семантичних веб-застосунків Jena [60]. SPARQL отримує вимоги застосування у формі запиту і повертає цю інформацію у вигляді набору зв'язків або графа RDF.

Тепер розглянемо існуючі рішення для інтеграції інформаційних ресурсів. Розпочнемо із засобів генерації і виконання семантичних запитів до розподілених джерел даних, які можуть надавати дані у форматах абстрактної моделі RDF.

Оброблювач ARQ запитів SPARQL для Jena надає розширення мови запитів для виконання запитів до розподілених джерел, розширюючи SPARQL для передавання шаблону триплету до віддаленої кінцевої точки доступу [60]. Це не вирішує підпроблему розподіленої обробки запитів у повному обсязі, оскільки клієнт повинен знати джерело, до якого необхідно направляти запит.

Низка проектів пропонує єдиний інтерфейс для виконання запитів до довільної кількості розподілених кінцевих точок і прозорий для клієнта поділ запиту на підзапити та їх виконання (федерування) [61]. Так, у побудованому на SPARQL-стандарті проекті DARQ сервіс-дескриптори використовуються для динамічного додавання і вилучення кінцевих точок, на основі статистичної інформації оптимізуються запити. У проекті SemWIQ підтримуються запити до розподілених джерел даних у розширеному SPARQL-синтаксисі, підтримується робота з багатьма кінцевими точками з використанням

конструкції COUNT [62]. Оптимізація здійснюється на основі статистичних даних про наповненість джерел даних в умовах наявності обмежень автономності, приватності та ін. У проекті FedX [63] запропоновано нові стратегії оброблення SPARQL-запитів до розподілених джерел даних на основі побудови єдиного об'єднаного RDF-графа. Якщо джерела даних надають свої SPARQL-кінцеві точки, немає потреби у внесенні змін у їх структури даних. У праці [64] розроблено адаптивний алгоритм виконання запитів на SPARQL-кінцевих точках, у якому статична оптимізація замінена динамічним об'єднанням результатів виконання підзапитів за наявності зв'язуючих значень.

Реальний проект PANGAEA [65, 66], присвячений створенню інформаційної системи за принципом бібліотеки відкритого доступу, забезпечує архівування, публікування і розповсюдження геоцентричних даних досліджень Землі. Система забезпечує довготерміновий доступ до контенту через інформаційні ресурси низки організацій, ідентифікацію, розділення, публікування і цитування даних, які архівуються як додатки до публікацій або цитовані колекції даних. Цитування доступне через портал German National Library of Science and Technology.

Перспективним видається підхід, побудований на концепції просторів даних. Запропонований А. Гелеві [67-69], він реалізований в ряді проектів [70-74], в Україні розвивається Н. Б. Шаховською [75-77]. За умови створення конструктивного формалізму як основи визначення, пошуку і опрацювання даних проміжного шару можна сподіватися, що користувачі отримають ефективну і зручну технологію інтеграції. На сьогодні це лише концепція, реалізація якої вимагає багато зусиль.

Тепер розглянемо поступи у моделях і технологіях реалізації мультиагентних систем, які можуть надати основу для зв'язування компонентів інтеграції, організації виконання запитів і швидкого створення ресурсів ССД. У найвідоміших методологіях TROPOS [78], Prometheus [79], ADELFE [80], Gaia [81] використовуються нотації агентів, починаючи з ранніх

стадій проектування, програмні каркаси, такі як JADE [82]. Вони мають багато спільного із сервісно-орієнтованою архітектурою [83] та багат шаровим програмним забезпеченням.

Загалом існуючі рішення з інтеграції базуються на технологіях створення, функціонування і розвитку розподілених систем. Незважаючи на відмінності різних технологій створення розподілених сервіс-орієнтованих систем, загальні принципи і теоретико-методологічні підходи завжди залишаються схожими. Різномірні сервіси необхідно внести в реєстр, описати і забезпечити можливість прозорого спілкування з клієнтами і між собою. Для цього звичайно використовують стек протоколів, схема взаємодії яких для організації сервісів наведена на рис. 1.5.

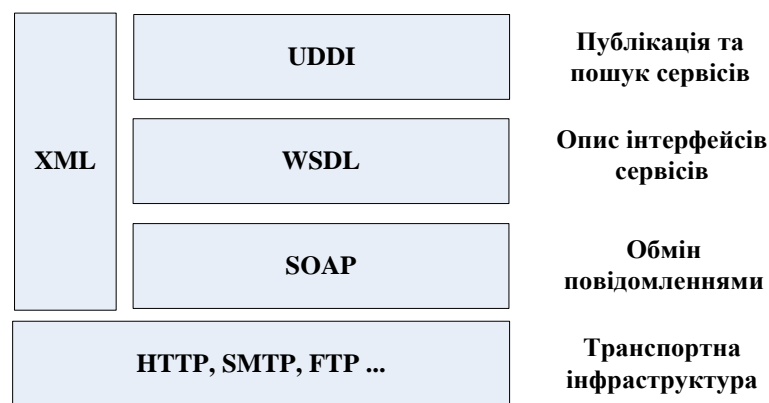


Рис. 1.5. Схема взаємодії протоколів для організації сервісів обслуговування

Для ведення реєстру сервісів використовується технологія Universal Description, Discovery and Integration (UDDI) [84], яка забезпечує публікацію інформації про сервіс і пошук потрібного сервісу людиною і програмою-клієнтом.

Для уніфікованого описання сервісів використовується WSDL – формат XML для описання мережних сервісів як набору операцій, які працюють за допомогою повідомлень з документо- або процедурно-орієнтованою інформацією [85]. Документи WSDL утворюють уніфікований прошарок, який уможливорює використання сервісів, створених за допомогою різних мов програмування, платформ, описують інтерфейс, URL, механізми комунікації, методи, які надає сервіс з відповідними параметрами (тип, назва, локація

Listener`а сервісу), структуру повідомлень сервісу. Формування WSDL-описів сервісів надає розробникам можливість працювати з сервісами як з звичайними класами [86].

Для обміну повідомленнями можна використовувати одну з поширених технологій, обґрунтовуючи вибір відповідністю вимогам створюваної платформи:

SOAP – проста і зручна у використанні разом з BPEL технологія, яка забезпечує взаємодію розподілених систем, незалежно від об'єктної моделі, операційної системи або мови програмування. Дані передаються у вигляді XML-документів особливого формату [87-89];

CORBA (Common Object Request Broker Architecture) – інструмент для розроблення і розгортання об'єктно-орієнтованих прикладних систем, який уможливорює взаємодію програм на різних мовах програмування, які виконуються в різних вузлах мережі. Це досягається за рахунок уніфікованого конструювання їх не залежних від операційної системи чи архітектури процесора інтерфейсів за допомогою декларативної мови Interface Definition Language (IDL) [90];

REST (Representational state transfer) – стиль архітектури програмних систем розподілених, який використовується для побудови веб-служб. Кожна одиниця інформації однозначно визначається глобальним ідентифікатором URL у строго заданому форматі. Дані передаються без будь-яких прошарків [91, 92];

WCF (Windows Communication Foundation) – платформа створення програм, які обмінюються даними у мережі не залежно від стилю і протоколу [93].

І хоч сьогодні найбільш прийнятною видається SOAP, на основі якої здійснюється взаємодія за схемою, наведеною на рис. 1.8, всі технології надають можливість інтегрувати сервіси різноманітного походження.

Для створення і підтримки рішень користувачів з врахуванням особливостей сервісного, процесно-функціонального і компонентно-

базованого підходів можна використовувати такі архітектурні способи організації групової функціональності:

- «хореографія» (Web Service Choreography), яка визначає протоколи взаємодії веб-сервісів, коли вони кооперуються для виконання однієї глобальної задачі, виконуючи окремі її частини. Ефективний для невеликих задач спосіб із зростанням складності задач швидко втрачає ефективність, оскільки зростає кількість задіяних сервісів і рішення стають надто громіздкими [94, 95];
- «оркестрація» (Service orchestration), при якій сервіси об'єднуються сервісом-окрестартором в один багатофункціональний бізнес-сервіс [96]. Роль окрестарторів можуть виконувати агенти, що дозволяє не змінювати структуру системи, комбінуючи переваги оркестрації і MAC. Для реалізації оркестрації розроблено мову для описання бізнес-процесів і протоколів їх взаємодії між собою BPEL (Business Process Execution Language), яка визначає координування оркестратором роботи підпорядкованих йому сервісів. Координована взаємодія досягається через обмін web-сервісів повідомленнями. Для довготривалих складно-структурованих бізнес-процесів важливі повідомлення, порядок обміну, стан, результати, час та інші аспекти бізнес-процесів. Роботу базових веб-сервісів організує агент як сервіс-оркестратор більш високого рівня, організаційні питання (послідовність обміну повідомленнями, врахування завантаженості та ін.) вирішують агенти рівнем вище. З BPEL застосовують SOAP і інші альтернативи [88, 89, 97].

На основі наведених технологій запропоновано декілька рішень, наприклад GEOSS (Global Earth Observation System of Systems), для накопичення, оброблення і обміну науковими даними. Вони досить поширені, хоч їм властивий ряд недоліків щодо оброблення різномірної інформації:

- складність об'єднання різномірних типів даних від різних джерел в одному запиті і розширення функціональності і збільшення степені

автоматизації, оскільки архітектура цих систем не передбачає можливості об'єднання в єдину мережу;

- відсутність єдиної моделі метаданих для всіх об'єктів метаданих і єдиного стандарту публікації, зберігання і оброблення даних;
- вимагання від користувача специфічного набору знань, вмінь і навичок для складання запиту і інтерпретації результатів його виконання.

Існує також Taverna Workflow Management System – система для розроблення, проектування і виконання складних потоків робіт, яка дозволяє координувати використовувати різноманітні сервіси WSDL, BioMoby, SoapLab та інших типів [98]. Однак, оскільки Taverna є науковим інструментом розроблення, їй бракує інтелектуальної складової і будь-яку робочу послідовність сервісів необхідно збирати вручну, що суттєво обмежує гнучкість і сферу застосування цієї системи [99].

Отже, жодне з цих рішень не забезпечує повну інтеграцію різноманітних джерел даних. Побудовані на цікавих ідеях, вони вирішують конкретні бізнесові або наукові проблеми. Вихідні коди цих рішень часто недоступні, в структурі їх даних важко розібратися навіть досвідченим користувачам. Ці проекти спрямовані на використання обмеженого набору моделей даних джерел. Оптимізовані під конкретну задачу, або навпаки, занадто уніфіковані, часто не розраховані на взаємодію з не-RDF даними, не спрямовані на структурування та класифікацію даних, що дуже важливо при роботі з інформаційними ресурсами Світової системи даних, вони не можуть в повному обсязі бути використані для вирішення поставленої в цій роботі проблеми. Потрібен більш гнучкий і універсальний підхід.

Але реалізація такого підходу вимагає розв'язання наведених вище задач. Розв'язання реальних проблем користувачів Світової системи даних вимагає інтеграції як джерел даних, так і застосунків. Розглянемо природу і особливості цієї домінуючої тенденції в ССД. Розпочнемо з аналізу реальних проектів Світової системи даних, реалізація яких вимагала інтеграції джерел і

застосунків. Це дозволить нам глибше зрозуміти потреби, які вимагають інтеграції, і власне процеси інтеграції різнорідних джерел даних і застосунків, виявити важливі їх особливості для досягнення сформульованої цілі роботи.

СЦД-Україна в рамках Державної цільової програми «Інформаційні та комунікаційні технології в освіті і науці» на 2006-2010 рр. виконував роботи, пов'язані з інтеграцією програмних і апаратних засобів партнерів в єдину інформаційну інфраструктуру, ядром якої стане розподілена інформаційна система обміну, оброблення і зберігання наукової інформації. Це дозволить користувачам отримати доступ до інформації з будь-яких доступних джерел незалежно від типу даних, носія інформації або технічних особливостей обладнання.

Така система, створювана за модульним принципом на базі відкритих стандартів, має широкі можливості інтеграції з інформаційними ресурсами Світової системи даних. Розроблення прототипів цієї системи показало перспективність переходу до платформи інтеграції [100].

СЦД з геоінформатики та сталого розвитку зі своїми партнерами розпочав ряд проектів, які дали певні результати. Розглянемо декілька найважливіших спільних проектів, реалізація яких вимагала інтеграції джерел даних і застосунків.

З іншого боку, досвід виконання СЦД-Україна ряду проектів з аналізу процесів сталого розвитку в глобальному і регіональному контексті, теж підкреслив потребу у використанні усіх наявних ресурсів, спонукав до пошуку нових форм співпраці і використання нових інструментів обміну даними, результатами роботи накопичених застосунків, породжував потребу в нових ресурсах. Крім того, реалізація зазначених проектів обумовила потребу в інструментах аналізу міждисциплінарних даних, побудованих на використанні відомих і розроблених нових методах штучного інтелекту.

В інститутах НАН України розпочали впровадження автоматизованої системи зберігання, передачі, оброблення і публікації, результати якої показали як потребу в розвинених моделях інтеграції даних, так створенні

набору моделей для вирішення тематичних задач з їх подальшою реалізацією на базі обчислювальних потужностей СЦД-Україна.

Розгортання на технічній базі СЦД-Україна та супровід електронної версії Національного атласу України (875 мап за розділами: загальна характеристика – 38 мап; історія – 79 мап; природні умови і природні ресурси – 320 мап; населення – 181 мапа; економіка – 181 мапа; екологічний стан навколишнього середовища) підтвердили необхідність стандартизації форматів із множини наявних, насамперед векторних isgeoMap, Adobe Flash та інших.

Був накопичений досвід створення інтерактивних ресурсів, доступ до яких надається на порталі СЦД-Україна (<http://wdc.org.ua/uk/services>):

1) Електронна версія Національного атласу України – інтерактивний ресурс, який містить 19 карт у форматі Adobe Flash, що дають уявлення про зміст кожного із 6 розділів. Структура, тексти до розділів, легенди цих 19 карт перекладено на англійською мовою;

2) Атлас сталого розвитку регіонів України – один з інструментів підтримки прийняття управлінських рішень на основі аналізу територіальних відмінностей в регіональному розвитку території України.

3) Розширена система публікації результатів моделювання процесів сталого розвитку – інформаційна система, яка забезпечує зручний і відкритий інтерфейс для загального доступу до даних зі сталого розвитку. Всі функції реалізовані у вигляді веб-сторінки і доступні цілодобово

4) Інтерактивні тематичні панелі WDC-Ukraine COVID-19 Dashboard для відстеження та прогнозування поширення COVID-19 в Україні – сервіс надає дані для України в цілому та у розрізі регіонів про кількість лабораторно підтверджених випадків COVID-19, летальних випадків і пацієнтів, які одужали. Інформацію подано у вигляді карт, діаграм, лічильників.

5) Короткостроковий прогноз COVID-19 – результати короткострокового прогнозного моделювання кількості хворих на COVID-19

в Україні і м. Києві, отримані з використанням багатошарової нейронної мережі Back Propagation на основі механізму «ковзного вікна».

б) Моніторинг динаміки поверхневих вод – серія інтерактивних тематичних панелей для аналізу динаміки поверхневих вод водосховищ Криму, Донецької та Луганської області.

Цей досвід показав, що назріла потреба в інтеграції, каталогізації і стандартизації наявних ресурсів, яка викликає необхідність у створенні необхідних моделей і методів. Була опрацьована розподілена модульна архітектура системи, яка забезпечує легке віддзеркалення даних і усунення вузьких місць мережі, пов'язане з трансконтинентальних розташуванням баз даних і є ключовим фактором для швидкої доставки і візуалізації даних. Розподілена мережа синхронних баз даних і серверів застосунків, реалізованих повністю на Java і доступних через Інтернет, продемонструвала перспективність інтеграції з ресурсами інших СЦД незалежно від використовуваних в них технологій. Стало зрозуміло, що для користувача важливо не тільки отримати дані по окремим країнам або регіонам, але і порівнювати їх в табличній і візуальній формі з метою аналізу. Підтверджена необхідність індексації наукових даних, пошуку за гнучкою системою критеріїв, високої швидкості оновлення пошукового індексу та індексації всіх поширених форматів – pdf, ppt, doc, xls, xml, rtf, txt та інших.

На основі огляду можна зробити обґрунтований висновок, що створення Світового центру даних в Україні не тільки сприяло поліпшенню ситуації з доступністю світових даних для українських дослідників, а й в значній мірі стимулювало і стимулює процес інтеграції української науки в світову.

Але на фоні змін в традиційних уявленнях про інформаційні процеси, ознаками якої стали хмарні обчислення і сервісний підхід, діючі СЦД ще не повною мірою інтегрувалися в сформоване ІТ-середовище з притаманними йому консолідацією засобів, конвергентністю сервісів, віртуалізацією ресурсів. Їх потужні системи накопичення і оброблення інформації виявилися не цілком адекватними потребам науки, недостатньо гнучкими для їх

використання в міждисциплінарних дослідженнях.

Тому постає завдання впровадити новий, координований глобальний підхід до організації збереження, оброблення і обміну науковими даними та інформацією, який гарантує універсальний рівноправний доступ до якісних даних та інформації для досліджень, освіти та прийняття рішень.

Виконання цього завдання породжує низку технічних задач, пов'язаних з уніфікацією форматів і протоколів передачі даних, організацією контролю якості наукових даних та інформації та ін. [2]. Але постає також задача організації єдиного інформаційного простору Українського сегменту і ССД та мережі партнерів ССД з геоінформатики та сталого розвитку.

Для вирішення цих завдань в ССД-Україна зроблено багато. Одночасно, виконаний аналіз показав, що проблема не може бути вирішена простим накопиченням нових інформаційних ресурсів, інструментів збору, накопичення, перетворення та публікування даних. Постає потреба у створенні цілісної платформи підтримки міждисциплінарних досліджень в ССД, що, в свою чергу, породжує необхідність розроблення ефективних моделей і методів інтеграції застосунків та інших складових підходу. Уточнимо задачу дослідження.

1.5. Уточнення задачі дослідження

В центрах Світової системи даних є багато формально не пов'язаних джерел даних, побудованих на базі різних інформаційних технологій – бази даних під управлінням СУБД, текстові файли, XML-документи, NoSQL-дані та інші. Виникає потреба, не вносячи зміни у структури джерел даних, що склалися історично, зробити їх доступними для користувачів за принципом «єдиного вікна». Одночасно необхідно забезпечити можливість виділення пов'язаних полів даних і виконання єдиного запиту до декількох джерел даних. Ще одним важливим аспектом проблеми є забезпечення можливості семантичного «розфарбування» існуючих даних, що зробить їх придатними

для машинного оброблення із застосуванням технологій семантичного рівня і більш ефективного пошуку.

Для вчених із різних галузей науки необхідно розробити і реалізувати зручні і одночасно ефективні механізми генерування проблемно-орієнтованих застосунків для аналітичного оброблення міждисциплінарних даних. Це, в свою чергу, вимагає створення комплексу моделей і методів описання джерел з урахуванням змісту даних, розбору запитів, їх федерування, планування і генерування власне застосунків.

При цьому необхідно створити математичні моделі і методи як основу цілісного рішення для забезпечення центрів збереження і оброблення даних, яке надасть гнучкі можливості на рівні інтеграції джерел даних і інформації і доступності сервісів для користувачів. Зазначені моделі і методи повинні бути позбавлені недоліків алгоритмічного підходу, пов'язаних з необхідністю перепрограмування алгоритмів оброблення даних при появі нових типів даних і зміні реалізованих чи появі нових алгоритмів. На рівні оброблення даних таке рішення повинне забезпечити:

- розподіленість обчислень і використання віддалених машинних ресурсів;
- гнучкість і адаптацію до навантаження на систему;
- простоту в використанні системи;
- надання даних за допомогою віддаленого клієнта або сервіса;
- доступність засобів інтелектуального оброблення даних безпосередньо кінцевому користувачу без спеціальних знань і навичок;
- швидку і просту інтеграцію даних і застосунків різноманітних глобальних інформаційних систем.

У найбільш загальному вигляді постає задача створення теоретичних основ і реалізації платформи підтримки міждисциплінарних досліджень в Світовій системі даних. Для побудови відповідного комплексу моделей і методів інтеграції джерел даних і застосунків, генерування проблемно-

орієнтованих застосунків для аналітичного оброблення міждисциплінарних даних, обґрунтування основних архітектурних рішень щодо платформи підтримки міждисциплінарних досліджень в Світовій системі даних, визначення структура і принципів функціонування та взаємодії її окремих компонентів і особливостей використання на прикладі побудови стратегій та сценаріїв сталого розвитку України необхідно виконати низку додаткових досліджень.

Запропоноване бачення загальної задачі створення платформи підтримки міждисциплінарних досліджень в Світовій системі даних треба трансформувати у фундаментальний науковий підхід, сформувавши теоретичні засади та окресливши інструментальні засоби, спрямовані на бажаний розвиток Світової системи даних. Додаткові дослідження треба виконати, щоб розробити, реалізувати та впровадити відповідні:

- моделі та алгоритми оброблення даних, спрямованих на систематичний аналіз поведінки складних антропогенних та природних систем;
- методи і технології моделювання складних антропогенних та глобальних природних систем;
- універсальні методи аналізу міждисциплінарних даних для використання в рамках інтегрованої системи доступу до інформаційних ресурсів Світових центрів даних в Україні;
- моделі та методи організації і виконання інтелектуального оброблення міждисциплінарних даних із наданням користувачам можливості формулювання запитів у підмножині природної мови;
- систему показників сталого розвитку регіонів України на основі комбінованого використання причинності та стохастичної семантики для використання в рамках інтегрованої системи доступу до інформаційних ресурсів Світових центрів даних в Україні;
- загальний підхід, моделі та методи систематичного коригування даних різної природи в інфраструктурі розподілених

міждисциплінарних баз даних українського сегменту Світової системи даних для вирішення фундаментальних міждисциплінарних завдань кореляції процесів різної природи.

Для втілення прийнятих ССД принципів розвитку, насамперед переходу від автономно функціонуючих СЦД та окремих служб до загальної глобально сумісної оперативно розподіленої системи даних, необхідно розробити, реалізувати і дослідити інструменти для:

- створення загального каталогу даних (єдиної точки доступу) для забезпечення доступу до неоднорідних джерел даних на основі агентських і онтологічних підходів. Доцільно розробити робочу версію каталогу і надати доступ для її тестування через Інтернет. Такий підхід дозволить об'єднати практично необмежену кількість різноманітних джерел даних в єдине неоднорідне середовище, яке буде прозорим для користувача та забезпечить організацію інтелектуального підбору наборів даних відповідно до запитів користувачів;
- розроблення ГІС для опрацювання при виконанні міждисциплінарних досліджень складних систем різної природи картографічних даних та відображення результатів в системі геоінформаційних ресурсів українського сегменту ССД;
- розроблення та дослідження засобів для збору та оброблення даних українського сегменту ССД.

За умови виконання визначених вище завдань в рамках ССД буде побудована гнучка масштабована платформа підтримки міждисциплінарних досліджень з доступом до даних та застосунків через Інтернет.

Висновки за розділом

1. Створення Українського сегменту ССД та його успішна діяльність стали важливим кроком у розвитку національної інфраструктури даних для наукових досліджень в Україні, що стало можливим завдяки реалізації проектів за фінансової підтримки Національної академії наук України, Міністерства освіти і науки України, Державного фонду фундаментальних досліджень, Національного фонду фундаментальних досліджень України тощо.

2. Показано, що діяльність ССД спрямована на вирішення завдань міждисциплінарного характеру, результати яких важливі для всіх їх партнерів. Такі дослідження ґрунтуються на взаємодії учасників на рівні засобів обміну і перетворення даних, технологій їх оброблення і аналізу, включаючи інструменти системного узгодження міждисциплінарних даних, їх систематизації, інтелектуального оброблення, оцінювання адекватності, аналізу якості, коректності тощо.

3. Розглянуті сутність і особливості організації і виконання міждисциплінарних досліджень в ССД, виявлені причини неефективного використання накопичених в ССД даних і обґрунтована необхідність побудови платформи підтримки міждисциплінарних досліджень в ССД, яка за рахунок драйверів розвитку інформаційних технологій нівелюватиме вплив зазначених причин і створюватиме умови розвитку міждисциплінарних досліджень.

4. Виконаний аналіз Світової системи даних і її діяльності як об'єкту управління, вимог до її доступності, оперативності, надійності, якості сервісів, на основі результатів якого обґрунтовано перспективність створення масштабованої платформи підтримки міждисциплінарних досліджень.

5. Показано, що проблема створення платформи підтримки міждисциплінарних досліджень в Світовій системі даних декомпонується на декілька підпроблем різного рівня складності, насамперед інтеграцію застосунків і джерел даних, узгодження даних, впровадження проблемно-

орієнтованого апарату інтелектуального аналізу даних та швидкого створення інформаційних ресурсів.

6. Наведено результати огляду існуючих рішень, які можна використати для підтримки міждисциплінарних досліджень СЦД ССД, встановлено, що жодне з них у повному обсязі не реалізує функцій платформи підтримки міждисциплінарних досліджень СЦД ССД, обґрунтовано перспективність створення вітчизняної платформи підтримки міждисциплінарних досліджень СЦД ССД і відібрано моделі, методи, інструменти і технології, які можна використати з цією метою.

7. Поставлена задача створення платформи підтримки міждисциплінарних досліджень в ССД і уточнена низка додаткових досліджень для побудови відповідного комплексу моделей і методів інтеграції джерел даних і застосунків, генерування проблемно-орієнтованих застосунків для аналітичного оброблення міждисциплінарних даних, обґрунтування основних архітектурних рішень щодо платформи підтримки міждисциплінарних досліджень в ССД, визначення структура і принципів функціонування та взаємодії її окремих компонентів і особливостей використання на прикладі побудови стратегій та сценаріїв сталого розвитку України та прогнозування пандемій.

РОЗДІЛ 2.

МОДЕЛІ І МЕТОДИ ІНТЕГРАЦІЇ ЗАСТОСУНКІВ І ДЖЕРЕЛ ДАНИХ

2.1. Концепція інтеграції застосунків і джерел даних в рамках платформи підтримки міждисциплінарних досліджень в Світовій системі даних

На сьогодні існує більш ніж 50 СЦД, які за більш ніж 50 років вибудували систему накопичення, аналізу, оброблення і міжнародного обміну даними. В потужних системах збереження даних СЦД накопичені великі обсяги астрономічних, геофізичних та інших наукових даних. На серверах СЦД здійснюється оброблення цих даних, для чого застосовується багато різноманітних прикладних застосунків, побудованих на різних технологіях.

Безумовно, з точки зору вчених, які працюють над вирішенням різноманітних ресурсовимогливих проблем, важливо мати доступ не до великої посіяної системи можливостей, а до цілісної сукупності джерел даних і застосунків, налаштованих на задоволення їх конкретних потреб з дружнім інтерфейсом, який не вимагає спеціальних знань в галузі ІТ. Однак, система накопичення, аналізу, оброблення і міжнародного обміну даними СЦД не забезпечує такого доступу. Крім того, вона не була розрахована на зростання рівня вимог наукової спільноти і є недостатньо гнучкою для використання в міждисциплінарних дослідженнях. Тому в 2008 р. була створена нова міждисциплінарна структура – Світова система даних (ССД, World Data System – WDS) для напрацювання і впровадження нового, координованого глобального підходу до наукових даних, який гарантує універсальний рівноправний доступ до якісних даних для досліджень, освіти і прийняття рішень. Новій структурі доведеться вирішувати накопичені проблеми, насамперед уніфікації форматів і протоколів передачі даних, забезпечення зручного доступу до даних, організації контролю якості наукових даних [2].

Труднощі, які при цьому виникають, пов'язані з тим, що існуючі СЦД накопичують і пропонують неоднорідні дані, визначені специфікою

відповідної галузі науки і країни. Можливості телекомунікаційних мереж постійно зростають, як і можливості сучасних комп'ютерів і сховищ даних, відкриваючи дорогу новим технологіям розподілених обчислень – Grid-системам і хмарним обчисленням. Об'єднання раніше виокремлених систем збору, збереження і оброблення даних СЦД на новій технологічній інтеграційній базі забезпечить значне збільшення їх сумарної ефективності. Створення такої системи надасть вченим зручний централізований доступ до раніше розрізнених ресурсів, одночасно полегшуючи і пришвидшуючи науково-дослідну роботу по всьому світу.

Найнагальнішою із проблем, які вимагають ефективних рішень, є проблема інтеграції. Сьогодні в багатьох СЦД використовуються успадковані технології, побудовані на програмних системах з жорсткою структурою. Більш нові рішення побудовані на клієнт-серверних і веб-орієнтованих технологіях. Сервісний підхід до оброблення і аналізу даних не набув необхідного поширення. Іноді відсутня можливість підключення геоінформаційних систем і навіть найпростіших загальнодоступних сервісів типу Google Maps. Якщо до цього додати відсутність єдиного стандарту, то необхідність інтеграції джерел даних, в якій забезпечуються інтеграція застосунків і інтеграція баз даних, можна вважати достатньо обґрунтованою.

При цьому формування сучасного ІТ-середовища передбачає інтеграцію у таких зрізах:

- технологічний: інтегровані дані і сервіси повинні базуватися на єдиному стеку протоколів взаємодії одне з іншим, використовувати спільні формати передачі даних для того щоб уможливити інтеграцію;
- змістовний: повинен підтримуватися єдиний формат описання даних в семантичній структурі джерел даних;
- функціональний: використання різномірних сервісів в єдиній структурі для розв'язання поставлених користувачем задач.

2.1.1. Загальний опис підходу до інтеграції застосунків

Принциповим положенням створення розподіленої системи є спосіб організації взаємодії сервісів. Із двох основних відомих способів взаємодії сервісів – оркестрації і хореографії – для Світової системи даних більш ефективним буде перший. Дійсно, оркестрація, яка групує базові сервіси в ієрархічно скомплексовані системи, підпорядковуючи їх адміністраторам – сервісам «окрестраторам», дозволяє об'єднати сервіси за зручними для СЦД ознаками (галузь науки, функціональність, регіональне розташування, тощо), а сервісам-окрестраторам надати повноваження відповідно до політик міжнародного обміну даними. Таким чином, можна вибудувати просту і ефективну систему, в якій пошук необхідного сервісу або побудова їх композиції для складних запитів, притаманних міждисциплінарним дослідженням, не буде вимагати великого часу, оскільки кожний сервіс-оркестратор має інформацію щодо функціональності підпорядкованих сервісів і крім того, вони можуть координувати свої можливості в процесі планування і реалізації запитів користувачів. Включення нових сервісів у систему також не буде становити особливих проблем, оскільки для цього потрібно буде лише скласти опис сервісу, внести його в реєстр і закріпити за певним сервісом-окрестратором [101].

Такий підхід буде працювати, коли користувач точно знає свої інформаційні потреби і має відповідні знання і навички для складання ланцюжків запитів до відомих оркестраторів. Але важливіше те, що підхід дозволить працювати з системою користувачам, які не мають можливості ініціювати сервіси, вказуючи порядок їх роботи і уточнюючи параметри виконання і взаємодії. Користувач повинен лише вміти формулювати свої потреби в термінах певної предметної області. Але тоді для об'єднання сервісів і організації їх взаємодії в потрібному порядку системі необхідна інтелектуальна складова. Для цього використовуються інтелектуальні агенти, які становлять ядро системи, реалізуючи логіку її функціонування, спрямовану на підтримку формування запитів, планування їх виконання і

організацію взаємодії базових сервісів. Включення інтелектуальних агентів в ієрархічну структуру як сервісів-оркестраторів приводить до формування дворівневої системи, нижній рівень якої формують сервіси, які розв'язують базові задачі, а верхній – інтелектуальні агенти, які оркеструють базові сервіси. Пов'язані кожний з кожним інтелектуальні агенти, використовуючи реєстри підпорядкованих їм базових сервісів і їх функціональність, за рахунок спілкування можуть реалізувати методи логічного виведення. Результатом виведення і буде композиція дій базових агентів, виконання якої дозволить отримати розв'язок поставленої користувачем задачі. Ця композиція дій або вивід передається на нижній рівень, де й здійснюється функціональна реалізація процесу розв'язання задачі. На цьому етапі контроль передається агентам нижнього рівня, які видають лише кінцевий результат для відправлення користувачу.

Власне виконання запиту користувача здійснюється після формування виводу, який містить вказівники на одне або декілька джерел даних, композицію з ідентифікаторів методів, які сервіси системи повинні виконати над цими даними. Деякі запити не вимагають попереднього оброблення даних і задовольняються методами сервісів безпосередньо після отримання інформації із джерела даних. Інші ж вимагають певної послідовності попередньо виконаних над даними дій, і, як наслідок, включення методів інших сервісів. Потрібні для цього залежності описуються в аксіомах, які вводяться в загальну базу знань при реєстрації відповідних сервісів в системі. Зазначені операції дозволяють отримати кінцевий продукт роботи системи – дані, які становлять розв'язок задачі користувача. Але початком активної роботи системи можна вважати надходження сформульованого користувачем запиту. Хоч, з огляду на вимоги до інтерфейсу системи, необхідно допомогти користувачеві скласти запит у знайомих йому термінах.

У рішенні, яке пропонується, для створення реєстру сервісів використовується UDDI, для уніфікованого їх опису – WSDL. Ключова ж частина – обмін повідомленнями реалізується за допомогою SOAP, а

відбувається згідно дерева вирішення задачі, відновленого механізмом відновлення дерева вирішення на основі виводу.

Для реалізації такої взаємодії сервісів у ССД найбільш прийнятним видається логічний підхід, який не тільки дозволяє вийти на рівень сформульованих вище вимог, але й позбавлений недоліків традиційного алгоритмічного підходу. Дійсно, необхідно лише розробити метод виведення і механізм відновлення дерева вирішення, які будуть підтримувати процеси формування запиту, планування його виконання, відновлювати схему розв'язання задачі і реалізувати її. Для створення і описання цих складових підходу найбільш прийнятним виглядає логічний підхід. А для реалізації логічного підходу необхідно:

- описати існуючі застосунки і їх функціональні можливості на формальній мові;
- сформулювати правила виведення;
- визначитися з методом виведення;
- розробити алгоритм роботи механізму побудови дерева виконання запиту користувача на основі виводу;
- реалізувати зазначені методи і механізми в агентах системи.

2.1.2. Загальний підхід до вирішення проблеми інтеграції джерел даних

Для описання специфікації семантики джерел даних обрані дескриптивні логіки. Як інструмент подання узагальнених специфікацій обрано онтологічні специфікації. Позначимо онтологічний клас і властивість в контексті семантики як С-клас і С-властивість відповідно.

Для реалізації системи інтеграції будемо застосовувати стек технологій Semantic Web – абстрактну модель даних у вигляді орієнтованих графів «суб'єкт-предикат-об'єкт» RDF, мову описання онтологій OWL і мову запитів до RDF-подібних даних SPARQL.

Для високорівневого опису доступних в системі предметних областей

необхідно створити центральну онтологію-класифікатор, яка складається з ієрархії С-класів і не залежить від джерел даних. Центральна онтологія може бути розширена іншими онтологіями такої ж структури або парами онтологія-відображення. Кожна така пара забезпечує підключення для конкретного джерела даних. Для цього онтологія описує структуру і зв'язки його С-класів і С-властивостей, а відображення вказує співвідношення його С-класів і С-властивостей до реальної структури джерела даних, а також параметри підключення до нього. Таким чином, запит до джерела даних може формуватися в термінах С-класів та С-властивостей.

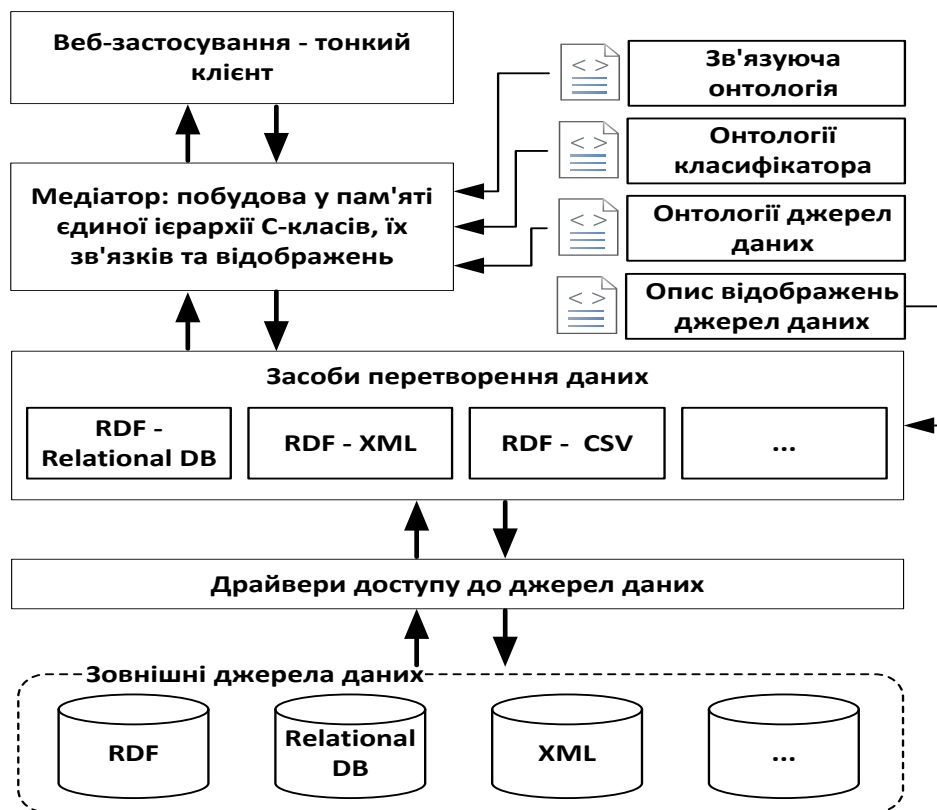


Рис. 2.2. Загальна схема рішення для підпроблеми інтеграції джерел даних

Оскільки різні джерела даних можуть мати спільні С-властивості, необхідно також створити онтологію для встановлення зв'язку між такими С-властивостями різних джерел даних за допомогою стандартних предикатів мови OWL.

На рис. 2.2 наведено загальну схему запропонованого рішення інтеграції джерел даних класу ЕІІ з посередником і web-орієнтованою архітектурою. На боці клієнта вибір даних в каталозі, формування запиту до джерел даних, а

також отримання відповіді на нього виконується за допомогою веб-застосунку. Медіатор є завантаженим в пам'ять сервера застосунком і містить побудоване дерево С-класів з онтологічних специфікацій (саме їх користувач бачить як каталог), їх зв'язки із відображеннями на реальні джерела даних, а також, опціонально, отримані з онтологічних специфікацій та опитування джерел даних обмеження на допустимі діапазони значень даних.

Зв'язування онтологій на рівні застосунку здійснюється за рахунок встановлення зв'язків їх С-класів з використанням унікального для кожного С-класу Uniform Resource Identifier (URI).

За прийнятими критеріями (гетерогенність, розподіленість і автономність) запропоноване рішення переважає рішення побудовані на множинних (multidatabase) і федеративних (federated database) базах даних, вже не кажучи про монолітні бази даних. Але за додатковими критеріями перевага не так відчутна.

Оброблення запитів до розподілених неоднорідних джерел даних. Концепція пропонованого рішення, схема оброблення запитів до розподілених джерел даних у якому наведена на рис. 2.3, позбавляє користувача від необхідності розуміння структур цих джерел, їх характеристик і особливостей мов запитів.

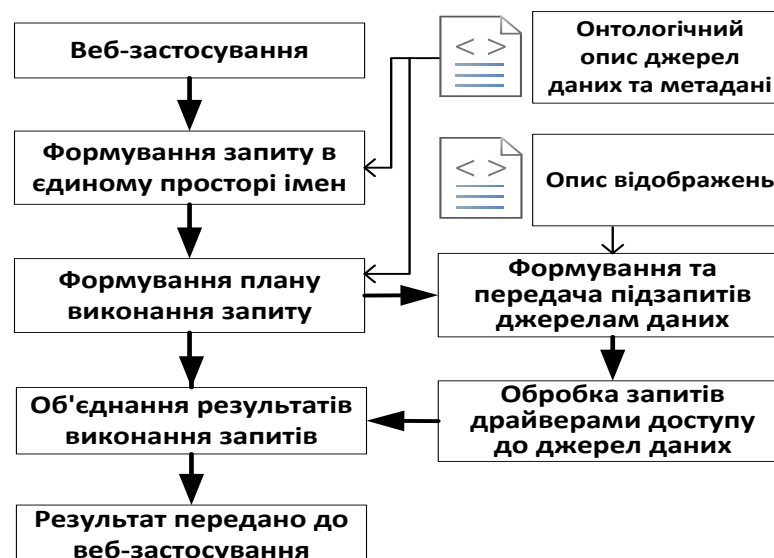


Рис. 2.3. Схема оброблення запитів у системі інтеграції джерел даних

Точкою взаємодії з користувачем є web-застосунок – тонкий клієнт. Він дозволяє користувачу формувати запит, оперуючи С-класами, С-властивостями, їх зв'язками та обмеженнями в глобальному просторі імен за допомогою меню-орієнтованого інтерфейсу. Інформація про структуру С-класів і С-властивостей, а також їх зв'язків завантажується з онтологічних описів джерел даних, а загальне дерево С-класів будується з онтології-класифікатора.

Далі застосунок формує план виконання запиту, визначаючи, які джерела даних можуть відповісти на запит або його частини, і чи можливо ці результати об'єднати. При цьому враховуються характеристики джерел даних, а також правила виведення, що визначають можливі операції над С-класами та їх С-властивостями, які отримані з онтологічних описів. Звідти ж береться інформація для формування умов-обмежень для джерел даних і запитів до них.

Після цього запит трансформується в єдиному просторі імен на основі описів відображень між джерелами даних. Він конвертується в запити до окремих джерел даних у відповідному синтаксисі, наприклад, у форматі SQL, і просторі імен.

Сформовані запити передаються у відповідні джерела даних за допомогою стандартних адаптерів, наприклад, JDBC. Оброблення запитів адаптером доступу до джерела даних є стандартною практикою при роботі з СУБД. Джерела даних функціонують автономно. Після отримання запиту вони його виконують і повертають результат у вигляді набору записів.

Отримавши від джерел даних результати виконання запитів, застосунки, дотримуючись сформованого плану виконання запиту, виконують необхідні операції – об'єднання, перетин, узагальнення, передача сервісам-обробникам тощо.

Виконана в дисертаційному дослідженні реалізація припускає, що всі джерела даних, застосовуючи стандартизовані адаптери, можуть обробляти SQL-або SPARQL-запити. Для взаємодії з джерелами даних, які надають XML-файли, або використовуються специфічні адаптери, або ці файли

приводяться у формат онтологій з метаданими, описаними у XML-подібному синтаксисі. Перший підхід вимагає налаштувань і вартісних програмних продуктів, а другий – втручання у структури джерел даних на боці провайдерів. Закладені рішення уможливають наступні реалізації, що будуть охоплювати більш широке коло джерел.

Реалізація запропонованої в роботі концепції інтеграції джерел даних базується на моделях визначення семантики предметної області. Розглянемо вимоги до цих моделей, які будуть наведені у розділі 3. Будемо виходити із проблем, вирішення яких вимагає створення інформаційної системи для інтеграції неоднорідних інформаційних ресурсів. Мова йде про такі підпроблеми загальної проблеми підтримки міждисциплінарних досліджень:

- забезпечення семантичної інтероперабельності інформаційних ресурсів;
- реалізація можливості порівняння вмісту цих ресурсів;
- встановлення відповідностей і запобігання конфліктів між інформаційними ресурсами;
- сполучення різнорідних ресурсів.

Як зазначалося вище, повна специфікація інформаційного ресурсу буде охоплювати специфікації власне джерела даних, обмежень цілісності і предметної області. Використання концепції медіатора вимагає описання структури і поведінка кожного інформаційного ресурсу.

Для описання інформаційних ресурсів будемо використовувати RDF і RDFS. Засновані на ідеях представлення знань, таких як семантичні мережі, фрейми і логіка предикатів, прості в реалізації вони дозволяють визначати класи, ресурси і властивості, ієрархії й обмеження типів, концепти, відношення й екземпляри. Для описання онтологій будемо використовувати мову OWL, яка повністю відповідає вимогам до розроблюваних систем інтеграції інформаційних ресурсів.

Концепція інтеграції передбачає планування виконання запитів та його оптимізацію. Автономність функціонування джерела даних ускладнює

отримання статистичних даних, але не звільнює від пошуку ефективних інструментів.

Щоб зосередитися на формулюванні вимог, критеріїв і обмежень до моделей і методів, які мають бути покладені в основу роботи компонентів, доцільно подати послідовність їх роботи:

- 1) генерування SPARQL-запитів;
- 2) парсинг, переписування і оптимізація запиту;
- 3) розбиття запиту на підзапити;
- 4) виконання підзапитів на вибраних джерелах даних;
- 5) очікування отримання відповідей від усіх джерел даних;
- 6) об'єднання отриманих результатів за допомогою Inner Join (опціонально) ;
- 7) повернення користувачу результату виконання запиту.

Тепер перейдемо до вимог. Формальні засоби метаопису джерел і формування запитів доцільно визначити на основі логічних, алгебраїчних і графічних формалізмів. Метаопис має зберігати необхідну для формулювання, парсингу, перетворення і оптимізації запиту інформацію. Для цього він має містити схему джерела (джерел) даних (наприклад, визначення таблиць, уявлень, типи даних і т.і.), схеми поділу (наприклад, інформацію про те, що глобальні таблиці були розділені і як вони можуть бути відновлені), фізичну інформацію (розташування розділів таблиць, структуру індексів і статистику для оцінювання ефективності виконання плану). Поданий у зручному для користувача вигляді запит має бути обробленим і переведеним у внутрішнє представлення, наприклад, граф запиту, зручне для парсингу.

Формування плану виконання запиту передбачає його еквівалентні перетворення, спрямовані на ліквідацію надлишкових предикатів, спрощення виразів і усунення вкладених запитів. План визначає послідовність виконання підзапитів до джерел даних та стратегію їх об'єднання. А для розподілених систем визначається, які частини запиту будуть виконуватися на яких джерелах даних.

Оптимізація плану враховує параметри (розмір таблиць, індекси, швидкість процесора тощо) і передбачає уточнення індексів і методів (наприклад, хешування і сортування), порядок застосування операцій (об'єднання і групування) і вузлів, на яких виконуються операції. План як послідовність виконання підзапитів до джерел даних і стратегія їх об'єднання доцільно подавати у вигляді дерева, вершини якого відповідають операторам (з'єднання, групування, сортування,...), а ребра вказують послідовність виконання операторів.

Оптимізація запитів у гетерогенній системі має бути універсальною і базуватися на оцінках «вартості» виконання запитів до джерел даних. Доцільно застосовувати такі стратегії оптимізації [38]:

- 1) Визначається узагальнена модель оцінювання вартості виконання запиту для всіх компонентів і її коефіцієнти корегуються для кожного компоненту за результатами виконання набору тестових запитів;
- 2) Визначається своя модель оцінювання вартості для кожного компоненту;
- 3) Вартість плану виконання запитів визначається на основі опрацювання статистичних даних моніторингу виконання компонентами запитів.

Найвищу точність оцінок вартості виконання запитів забезпечує другий підхід, але його реалізація найвитратніша.

Задача трансформації плану виконання запиту у виконуваний план може включати генерацію асемблероподібного коду для ефективного виконання окремих частин плану, наприклад предикатів і виразів, на джерелі даних.

Задача власне виконання запиту полягає у використанні загальних реалізацій для кожного оператора. Організація роботи сучасних підсистем виконання запитів базується на моделі ітератора, яка забезпечує реалізацію операторів і стандартний інтерфейс. Виходячи із необхідності збереження дисциплінарності СЦД, для інтеграції їх різнорідних джерел даних найперспективнішим видається підхід, реалізація якого дозволить звертатися

до даних, трансформуючи їх у сумісні формати у реальному часі і не зберігаючи в проміжних сховищах. Це дозволить доповнити діяльність СЦД широкими можливостями обміну даними і доступу кінцевих користувачів до ресурсів усіх СЦД, незалежно від їх розташування, форматів даних і програмних засобів аналізу і оброблення даних, що складе надійну основу для розвитку міждисциплінарних досліджень.

2.1.3. Мультиагентна взаємодія як основа інтеграції застосунків і джерел даних платформи підтримки міждисциплінарних досліджень

На сьогоднішній день у ССД накопичено велику кількість даних, що представляють інтерес для наукової спільноти, частина з яких семантично тотожні, але різнорідно представлені. Для ефективного використання цих даних необхідно забезпечити централізований спосіб доступу і механізм пошуку, який дозволить перетворювати різнорідні за поданням, але тотожні за семантикою, дані в метадані. Метадані є поданням деяких закономірностей в конкретній предметній області і придатні для подальшого об'єднання різнорідних вихідних даних. Проблема створення програмних засобів перетворення інформації в метадані особливо актуальна для міждисциплінарних досліджень в ССД.

СЦД мають власні структури зберігання даних, задані особливостями програмного і апаратного забезпечення, яке використовується для збору і оброблення інформації. Але дані не мають загального формату фізичного представлення і визначення семантики.

Будувати систему, здатну визначити семантично однорідні дані для подальшої маніпуляції ними, доцільно на основі концепції Linked Data [102]. Опис предметної області на RDF і для встановлення взаємозв'язків між об'єктами різних інформаційних ресурсів через URI дозволили створити ефективний інструментарій на web-порталах федеральних даних США www.data.gov, Німецької національної бібліотеки www.d-nb.de та інших організацій.

Але кожен з них має власну предметну область і єдине джерело даних. Пошук і семантизація інформації в системі, що складається з безлічі вузлів-джерел даних різної фізичної структури і належать до декількох предметних областей, не забезпечуються. Для усунення цих недоліків перспективним є використання мультиагентних систем (МАС) [103].

Щоб інтегрувати джерела даних в єдину систему, здатну забезпечити користувача необхідними метаданими і, зокрема, надати можливість визначити семантично однорідні дані для подальшої їх інтеграції та застосування до них різних методів перетворення, оброблення та подання, необхідно розробити архітектуру МАС для перетворення неоднорідної, але семантично тотожної інформації, в метадані, придатні для подальшого аналізу і дослідження.

В роботі запропонована архітектура МАС схематично представлена на рис. 2.4. На основі підходів, запропонованих в [103], в ній реалізоване середовище, яке інтегрує розподілені джерела даних, забезпечує дані структурами метаданих, визначеними внаслідок їх різноманітності на окремому від джерел даних рівні.

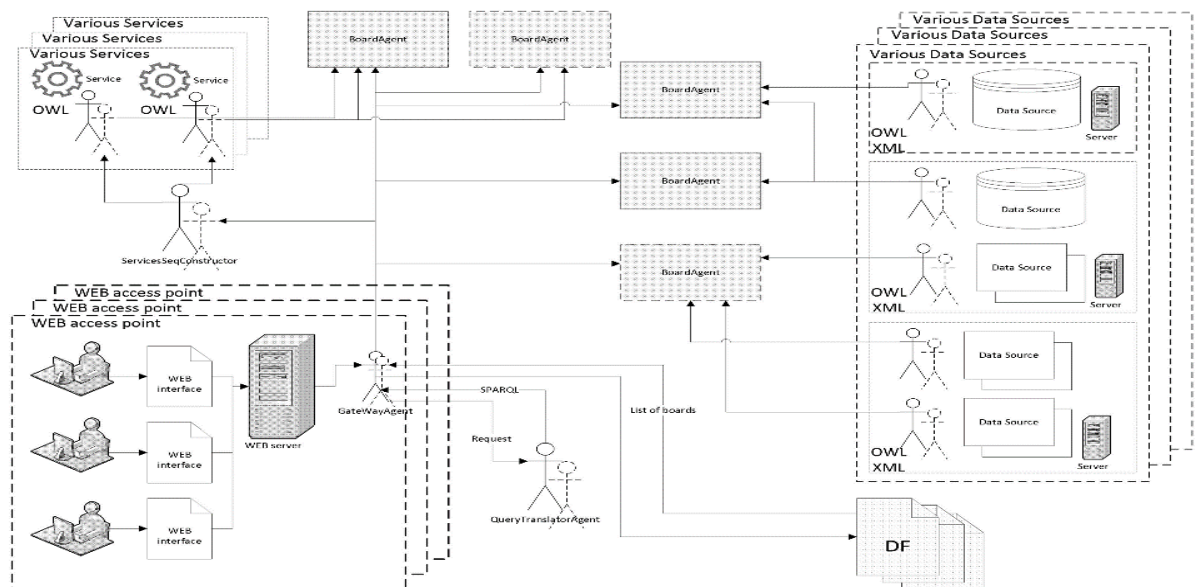


Рис.2.4. Архітектура МАС для інтеграції гетерогенних розподілених джерел даних

У розділі 3 буде обране середовище реалізації МАС, визначені необхідні для реалізації типи агентів, насамперед для повідомлень про нові запити, організації зв'язку, виконання і формування результатів, перетворення отриманих результатів. Також буде уточнено інтерфейс взаємодії з системою, спосіб звертання до джерел, можливість міграції, спосіб пошук агентів всередині МАС і сервісів.

Використання семантики даних, обґрунтований вище вибір способу описання семантики даних і парадигми забезпечення інтероперабельності даних і можливості виділення семантично однорідної інформації, інструментів проектування компонентів, зв'язку з джерелами даних, мови формулювання запитів користувача вимагають розв'язання складних задач реалізації архітектури МАС.

Існуючі на сьогодні проекти впроваджують семантизацію даних, але спеціалізуються на окремій предметній області і одному джерелі даних. Підхід, що пропонується, забезпечує можливість інтеграції множини джерел, які необхідно використовувати разом для вирішення міждисциплінарних наукових задач.

Реалізація підходу вимагає формалізації семантичного опису даних предметних областей, наприклад фізики твердої Землі, океанології та ін. Такий підхід може бути реалізований за умови використання механізму доступу до різномірних джерел даних на основі агентно-орієнтованого підходу. Моделі і технології інтеграції агентів з онтологіями предметних областей і джерел даних, на основі яких реалізована запропонована вище архітектура інтеграції гетерогенних розподілених джерел даних на основі агентно-орієнтованого підходу, будуть розглянуті у наступному розділі.

2.2. Логічна модель як формальна основа інтеграції застосунків

Ефективне використання переваг розподілених систем і можливостей сервіс-орієнтованих технологій вимагають забезпечення уніфікованого доступу до сервісів. Цьому сприяє формування нового демократичного ІТ-середовища [34].

У ССД накопичено багато даних і застосунків, які можна використовувати для вирішення актуальних проблем розвитку суспільства. Але цьому перешкоджають труднощі у взаємодії успадкованих застосунків, обумовлені відмінністю архітектур, різноманіттям форматів представлення даних та іншими чинниками.

Виникає потреба в інтеграції ресурсів із різних джерел, щоб вчені різних галузей знання могли використовувати накопичені ресурси, орієнтуючись на знання предметної області, а не особливості ІТ. Зазвичай інтеграцію даних в інформаційних системах розуміють як забезпечення єдиного уніфікованого інтерфейсу для доступу до певної сукупності, у загальному випадку, неоднорідних незалежних джерел даних [34, 37, 104]. Для користувача інформаційні ресурси всієї сукупності джерел, які інтегруються, повинні представлятися як нове єдине джерело, а система, яка забезпечує такі можливості, називається системою інтеграції даних [2]. Взаємодія застосунків дуже важлива для інтеграції нового ІТ-середовища. В умовах сервісного підходу інтеграцію застосунків можна розглядати як теоретико-методологічну основу впровадження компонентно-базованого підходу до створення інформаційних систем. Інтеграція даних в інформаційних системах і інтеграція як взаємодія застосунків мають комплексуватися, що дозволить:

- створювати механізми ефективного управління інформаційними ресурсами;
- розширювати сфери стандартизації протоколів і форматів обміну даними на основі взаємодії відкритих систем;
- створювати крос-платформні застосунки, інтегрувати дані і застосунки;

- спрощувати розгортання і розвиток систем за рахунок приховування внутрішніх деталей апаратного і програмного забезпечення.

Серед невирішених проблем залишаються проблеми створенню математичних моделей, методів і засобів, необхідних для інтеграції застосунків. У роботі пропонується підхід до вирішення зазначеної проблеми із застосуванням апарату математичної логіки.

Уточнимо постановку проблеми інтеграції застосунків. Необхідно створити математичні моделі і методи як основу цілісного рішення для забезпечення центрів збереження і оброблення даних, яке надасть гнучкі можливості на рівні інтеграції інформації і доступності сервісів для користувачів. Зазначені моделі і методи повинні бути позбавлені недоліків алгоритмічного підходу, пов'язаних з необхідністю перепрограмування алгоритмів оброблення даних при появі нових типів даних і зміні реалізованих чи появі нових алгоритмів. На рівні оброблення даних таке рішення повинне забезпечити:

- розподіленість обчислень і використання віддалених машинних ресурсів;
- гнучкість і адаптацію до навантаження на систему;
- простоту в використанні системи;
- надання даних за допомогою віддаленого клієнта або сервісу;
- доступність засобів інтелектуального оброблення даних безпосередньо кінцевому користувачу без спеціальних знань і навичок;
- швидку і просту інтеграцію даних і застосунків різноманітних глобальних інформаційних систем.

Розглянемо особливості інтеграції застосунків у СІД. На серверах СІД здійснюється оброблення даних, використовуються прикладні застосунки, побудовані на різних технологіях. Накопичуються і оброблюються неоднорідні дані різних галузей. Технічні можливості (мереж, комп'ютерів і сховищ) постійно зростають, відкриваючи дорогу новим технологіям – Grid-

системам і хмарним обчисленням [105]. Об'єднання раніше виокремлених систем збору, збереження і оброблення даних СЦД на новій технологічній інтеграційній базі забезпечить значне збільшення їх сумарної ефективності. Створення такої системи надасть вченим зручний централізований доступ до раніше розрізнених ресурсів, одночасно полегшуючи і пришвидшуючи науково-дослідну роботу по всьому світу.

У багатьох СЦД використовуються успадковані технології, побудовані на програмних системах з жорсткою структурою. Більш нові рішення побудовані на клієнт-серверних і веб-орієнтованих технологіях. Сервісний підхід до оброблення і аналізу даних поширюється. Формування сучасного ІТ-середовища передбачає інтеграцію у таких зрізах:

- технологічний: інтегровані дані і сервіси повинні базуватися на єдиному стеку протоколів взаємодії одне з іншим, використовувати спільні формати передачі даних для того щоб уможливити інтеграцію;
- змістовний: повинен підтримуватися єдиний формат описання даних в семантичній структурі джерел даних;
- функціональний: використання різнорідних сервісів в єдиній структурі для розв'язання поставлених користувачем задач.

Існуючі рішення базуються на технологіях створення, функціонування і розвитку розподілених систем. Незважаючи на відмінності різних технологій створення розподілених сервіс-орієнтованих систем, загальні принципи і теоретико-методологічні підходи завжди залишаються схожими. Різнорідні сервіси необхідно внести в реєстр, описати і забезпечити можливість прозорого спілкування з клієнтами і між собою. Для цього звичайно використовують стек протоколів, схема взаємодії яких для організації сервісів наведена вище. Необхідно лише розробити ефективні моделі і методи інтеграції застосунків.

Загальний підхід до вирішення проблеми. Як зазначено вище для створення розподіленої системи вибрано оркестрацію як спосіб організації

взаємодії сервісів. Реалізація базується на UDDI, WSDL, SOAP, BPEL і JADE (див. рис 2.1). Для організації взаємодії сервісів у ССД найбільш прийнятним є логічний підхід, який дозволяє вийти на рівень сформульованих вище вимог, позбавляє недоліків алгоритмічного підходу.

Запит користувача обробляється з урахуванням сервісів системи, їх функціональності (описаної аксіомами системи) і правил виведення, визначених логічним формалізмом. Отриманий вивід передається на нижній рівень для реалізації. На рівні сервісів, кожний агент виконує виклики необхідних сервісів із свого набору, і відправляє оброблені дані у відповідності з ланцюжком дій. Це відбувається згідно дерева вирішення задачі, відновленого механізмом відновлення дерева вирішення на основі виводу.

Дійсно, потрібно лише розробити метод виведення і механізм відновлення дерева вирішення, як основи формування запиту, планування його виконання, відновлювати схему розв'язання задачі і реалізувати її. Для реалізації логічного підходу необхідно:

- описати існуючі застосунки і їх функціональні можливості на формальній мові;
- сформулювати правила виведення;
- визначитися з методом виведення;
- розробити алгоритм роботи механізму побудови дерева виконання запиту користувача на основі виводу;
- реалізувати зазначені методи і механізми в агентах системи.

Розглянемо ці складові детальніше.

Формальна логічна система. Опишемо формалізм, на базі якого буде побудована програмна система, яка забезпечить вирішення поставленої проблеми. За основу візьмемо клаузальну логіку першого порядку і опишемо мову формальної системи, дотримуючись запропонованих у праці [3] визначень її конструктивних елементів.

Символи:

службові: $(,), [C,], \{, \}, :, <, >, ;$;

константи:

1) *індивідні*, простих типів (int, real, char, bool) – $a_1^1, a_2^1, \dots, a_1^2, a_2^2, \dots$ причому кожна константа a_i^k належить сорту (простому типу) k ; структурного типу (construct) – c_1, c_2, \dots ; процедурного типу (method) – d_1, d_2, \dots ; об'єктного типу (problem, entity, relation) – e_1, e_2, \dots ;

2) *функціональні i-місні*, для індивідів типу k – $h_1^1, h_2^1, \dots, h_1^2, h_2^2, \dots$;

3) *предикатні i-місні*, для індивідів типу k – $A_1^1, A_2^1, \dots, A_1^2, A_2^2, \dots$ (в цей клас включаються таксономічні, реляційні та інші предикати, а також традиційні відношення, щонайменше рівності $=$ і порядку \geq);

змінні: для індивідів типу k – $x_1^1, x_2^1, \dots, x_1^2, x_2^2, \dots$, причому кожна змінна x_i^k належить типу k ;

логічні: $\neg, \wedge, \vee, \leftarrow, \exists, \forall, \Leftrightarrow$.

Індивідні терми типу k :

- 1) кожна індивідна константа a_i^k типу k є індивідним термом типу k ;
- 2) кожна вільна змінна x_i^k для індивідів типу k є індивідним термом типу k ;
- 3) якщо h_i^j – деяка функціональна константа для індивідів типу k , τ_1, \dots, τ_j – терми для індивідів типу k , то $h_i^j(\tau_1, \dots, \tau_j)$ – є індивідним термом типу k ;
- 4) інших індивідних термів типу k немає.

Терми, отримані застосуванням правил побудови 1 або 2 визначення, назовемо простими, а всі інші – складними.

Формули для індивідів:

- 1) якщо A_i^j – предикатна константа для індивідів, τ_1, \dots, τ_j – терми для них, то $A_i^j(\tau_1, \dots, \tau_j)$ – атомарна формула для індивідів;
- 2) атомарна формула для індивідів – формула для індивідів;
- 3) інших формул для індивідів немає.

Надалі вважаємо, що в системі є універсальний перетворювач формул до традиційного для клаузуальної форми (мова йде про клаузи Хорна) вигляду з єдиним символом \rightarrow , атомарними формулами ліворуч і праворуч від нього, неявним квантором \forall .

Специфікатори – це конструкції вигляду τ_1 , де τ_1 – терм для індивідного об'єкту. Визначимо специфікатори конструктів:

1) якщо $e_1^e \dots e_i^e$, – індивідні терми типу entity, $e_1^r \dots e_i^r$, – індивідні терми типу relation, $A_i^j(a_1^1 \dots a_j^1) \dots A_i^j(a_1^k \dots a_j^k)$ – атомарні формули для індивідів простих типів, τ – індивідний терм типу construct, то $\tau: (e_1^e \dots e_i^e, e_1^r \dots e_i^r, A_i^j(a_1^1 \dots a_j^1) \dots A_i^j(a_1^k \dots a_j^k))$ – специфікатор конструкта.

2) інших специфікаторів конструктів немає.

Визначимо передумови:

1) якщо c_1 – специфікатор конструкта, Π – послідовність атомарних формул, то $\langle \tau_1: (c_1, \Pi) \rangle$ – елементарна передумова;

2) елементарна передумова – передумова;

3) якщо $\langle \tau_1 \rangle$ – передумова, а τ_2 – елементарна передумова, то $\langle \tau_1, \tau_2 \rangle$ – передумова;

4) інших передумов немає.

Визначимо постумови:

1) якщо τ_1 – специфікатор конструкта, Π – послідовність атомарних формул, то $\langle \tau_1: (c_1, \Pi) \rangle$ – елементарна постумова;

2) елементарна постумова – постумова;

3) якщо $\langle \tau_1 \rangle$ – постумова, а τ_2 – елементарна постумова, то $\langle \tau_1, \tau_2 \rangle$ – постумова;

4) інших постумов немає.

Визначимо специфікатори методів:

1) якщо τ – індивідний терм типу method, $\langle \tau_1 \rangle$ – передумова, $\langle \tau_2 \rangle$ – постумова, то $\tau: (\langle \tau_1 \rangle, \langle \tau_2 \rangle)$ – специфікатор методу;

- 2) арність по конструктам передумови методу має бути не меншою арності по конструктам постумови методу;
- 3) інших специфікаторів методів немає.

Визначимо специфікатори проблем:

- 1) якщо $\langle \tau_1 \rangle$ – передумова, а $\langle \tau_2 \rangle$ – постумова, τ – терм для індивідного об'єкту типу Problem, то $\tau: \langle \langle \tau_1 \rangle, \langle \tau_2 \rangle \rangle$ – специфікатор проблеми;
- 2) інших специфікаторів проблем немає.

Клауза – це вираз вигляду $\Pi \rightarrow \Lambda$, де Π – послідовність атомарних формул; Λ – єдина атомарна формула. У залежності від вигляду атомарних формул будемо виділяти клаузи:

- 1) індивідні (містять тільки атомарні формули для індивідів);
- 2) типізовані (містять тільки атомарні формули для типів);
- 3) загального вигляду (містять атомарні формули для індивідів і типів).

Базу знань системи складають онтологія аксіом, які зображують методи сервісів системи, і онтології джерел даних, описаних на мові OWL на базі RDF. Також до бази знань належать правила виведення, необхідні для отримання потрібного результату у випадках, коли необхідно комбінувати методи як одного, так і різних сервісів.

Правила виведення:

1. Якщо $d_1: (\langle \tau_1 \rangle, \langle \tau_2 \rangle)$ і $d_2: (\langle \tau_3 \rangle, \langle \tau_1 \rangle)$, то $d_1: (\langle d_2 \rangle, \langle \tau_2 \rangle)$
2. Якщо $d_1: (\langle \tau_1 \rangle, \langle \tau_2 \rangle)$ і $d_2: (\langle \tau_3 \wedge \tau_4 \rangle, \langle \tau_1 \rangle)$, то $d_1: (\langle d_2 \rangle, \langle \tau_2 \rangle)$
3. Якщо $d_1: (\langle \tau_1 \wedge \tau_2 \rangle, \langle \tau_3 \rangle)$ і $d_2: (\langle \tau_4 \rangle, \langle \tau_1 \rangle)$ і $d_3: (\langle \tau_5 \rangle, \langle \tau_2 \rangle)$, то $d_1: (\langle d_2 \wedge d_3 \rangle, \langle \tau_3 \rangle)$
4. Якщо $d_1: (\langle \tau_1 \rangle, \langle \tau_2 \rangle)$ і $d_2: (\langle \tau_3 \vee \tau_4 \rangle, \langle \tau_1 \rangle)$, то $d_1: (\langle d_2 \rangle, \langle \tau_2 \rangle)$
5. Якщо $d_1: (\langle \tau_2 \wedge \tau_2 \rangle, \langle \tau_1 \rangle)$, то $d_1: (\langle \tau_2 \rangle, \langle \tau_1 \rangle)$
6. Якщо $d_1: (\langle \tau_1, \tau_2 \rangle, \langle \tau_3 \rangle)$ і $d_2: (\langle \tau_4 \rangle, \langle \tau_1 \rangle)$ і $d_3: (\langle \tau_5 \rangle, \langle \tau_2 \rangle)$, то $d_1: (\langle d_2, d_3 \rangle, \langle \tau_3 \rangle)$
7. Якщо $d_1: (\langle \tau_1 \rangle, \langle \tau_2, \tau_3 \rangle)$ і $d_2: (\langle \tau_2 \rangle, \langle \tau_4 \rangle)$ і $d_3: (\langle \tau_3 \rangle, \langle \tau_5 \rangle)$, то $d_2: (\langle d_1 \rangle, \langle \tau_4 \rangle)$ и $d_3: (\langle d_1 \rangle, \langle \tau_5 \rangle)$

2.3. Метод виведення в клаузуальній логіці взаємодії застосунків

Будемо використовувати запропонований у працях [3, 5, 20] метод, побудований на типізації тверджень і аналогії. Його ідея полягає у побудові виводу в абстрактному просторі з наступним його використанням для управління процесом виведення у вихідному просторі пошуку розв'язків. Це дозволить підвищити ефективність виведення за рахунок відсікання більшої частини безперспективних гілок виводу у вихідному просторі [106].

Надалі під мультіклаузою (m -клаузою) будемо розуміти мультимножину літералів (атомарних формул) і будемо записувати в m -клаузі літерал L стільки разів, скільки разів він повторюється. Звичайну клаузу будемо вважати m -клаузою, кратність кожного літерала в якій дорівнює одиниці. На m -клаузи переносяться операції \cup (об'єднання), \cap (перетин), $-$ (різниця), \cdot (конкатенація) і відношення \subseteq (входження) для мультимножин.

Нехай $A_1 \in C_1$, $A_2 \in C_2$ а α_1 і α_2 – такі підстановки, що для деякого літерала L у правій частині A_1 і лівій частині A_2 , $A_1\alpha_1 = \{L\}$, $A_2\alpha_2 = \{L\}$, причому, якщо $|A_i| > 1$, $i = 1, 2$, то відповідна підстановка α_i є найбільш загальним уніфікатором літералів із A_i . Тоді клауза, отримана з об'єднання клауз $C_1\alpha_1$ і $C_2\alpha_2$ вилученням L з правої і лівої частин називається m -резольвентою m -клауз C_1 і C_2 . Нехай всі m -клаузи вихідної множини S упорядковані, а C_1 , і C_2 дві з них. Нехай $A_1 \in C_1$, $A_2 \in C_2$ а α_1 і α_2 – такі підстановки, що для деякого літерала L у правій частині A_1 і лівій частині A_2 , $A_1\alpha_1 = \{L\}$, $A_2\alpha_2 = \{L\}$, причому, якщо $|A_i| > 1$, $i = 1, 2$, то відповідна підстановка α_i є найбільш загальним уніфікатором літералів із A_i . Тоді клауза, отримана з об'єднання клауз $C_1\alpha_1$ і $C_2\alpha_2$ вилученням L з правої частини і останнього L лівої частини і викреслюванням будь-якого не підкресленого літерала, за яким немає жодного іншого літерала, називається упорядкованою лінійною m -резольвентою упорядкованих m -клауз C_1 і C_2 . У випадку, коли останній літерал у лівій частині упорядкованої m -клаузи уніфікується з підкресленим літералом у правій частині цієї ж клаузи, упорядкована лінійна m -резольвента може бути отримана вилученням

останнього літерала лівої частини упорядкованої m -клаузи, тобто редукцією m -клаузи.

Назвемо m -резолюційним виводом Tm пару, перший компонент якої – множина вершин виводу, а другий – множина трійок вершин (для виділення компонентів будемо використовувати пару функцій-селекторів: $s-N(Tm)$ і $s-M(Tm)$ відповідно). Кожна вершина m -резолюційного виводу характеризується позначкою і значенням глибини у виводі таким чином, що якщо $n \in s-N(Tm)$, то $s-L(n)$ – позначка вершини n і $s-D(n)$ – значення її глибини у виводі. Якщо $\langle n_1, n_2, n_3 \rangle \in s-M(Tm)$, то $s-L(n_3)$ є m -резолювентою $s-L(n_1)$ і $s-L(n_2)$. Кожна трійка такого вигляду називається m -резолюцією. Якщо $n \in s-N(Tm)$ і вершина n не є третьою складовою жодної з трійок з $s-M(Tm)$, то n називається початковою вершиною виводу Tm . Позначка такої вершини – початкова m -клауза. Якщо $n \in s-N(Tm)$ і вершина n не є ні першою, ні другою складовою жодної з трійок з $s-M(Tm)$, то n називається термінальною вершиною виводу, а її позначка – термінальною m -клаузою. Загалом Tm має задовольняти таке обмеження: якщо $\langle n_1, n_2, n_3 \rangle \in s-M(Tm)$, то $\langle n_2, n_1, n_3 \rangle \in s-M(Tm)$. Значення глибини (в виводі) кожної вершини n , $n \in s-N(Tm)$, за визначенням: 1) $s-D(n) = 0$ для будь-якої початкової вершини n ; 2) $s-D(n) = 1 + \min\{\max\{s-D(n_1), s-D(n_2)\} \mid \langle n_1, n_2, n \rangle \in s-M(Tm)\}$ для будь-якої непочаткової вершини n .

Назвемо m -резолюційний вивід Tm виводом із S , якщо позначки початкових вершин Tm належать множині m -клауз S . Із S виводиться C , якщо Tm є виводом із S , а C – позначкою однієї з вершин Tm . Нехай значення функції $\text{Result}()$ визначене для Tm і дорівнює C , якщо термінальна клауза C виводу Tm є єдиною, тобто $\text{Result}(Tm) = C$. Нехай Tm_1 і Tm_2 – m -резолюційні виводи. Назвемо Tm_1 субвиводом Tm_2 (позначимо $Tm_1 \subseteq Tm_2$), якщо виконуються такі умови: $s-N(Tm_1) \subseteq s-N(Tm_2)$ і $s-M(Tm_1) \subseteq s-M(Tm_2)$. При цьому, якщо всі початкові вершини Tm_1 – початкові вершини Tm_2 , то Tm_1 назвемо початковим субвиводом Tm_2 .

Нехай Tm – m -резолюційний вивід із S , всі m -клаузи якого упорядковані. Якщо для будь-якої трійки $\langle n_1, n_2, n_3 \rangle$ із $s-M(Tm)$ $s-L(n_3)$ є упорядкованою лінійною m -резольвентою $s-L(n_1)$ і $s-L(n_2)$, то Tm назвемо упорядкованим лінійним виводом із S . Тоді, за визначенням m -резолюційного виводу: $\max\{s-D(n_1), s-D(n_2)\} = s-D(n_3) - 1$.

Крім загальних властивостей m -резолюційних виводів, упорядковані лінійні m -резолюційні виводи мають особливості. Так, для упорядкованого лінійного m -резолюційного виводу Tm справедливі такі умови: 1) якщо $\langle n_1, n_2, n_3 \rangle \subseteq s-M(Tm)$, то $\langle n_2, n_1, n_3 \rangle \notin s-M(Tm)$; 2) для кожної непочаткової вершини n_3 глибини r існує трійка $\langle n_1, n_2, n_3 \rangle$, в якій n_1 має глибину $r-1$, n_2 відповідає початковій вершині, або відсутня; 3) Tm містить єдину термінальну клаузу.

Для управління упорядкованим лінійним виводом у праці [3] використовується абстракція типізації. Нехай f – відображення із множини m -клауз в множину підмножин m -клауз таке, що: 1) якщо m -клауза C_3 є m -резольвентою m -клауз C_1 і C_2 , а $D_3 \in f(C_3)$, то існують $D_1 \in f(C_1)$ і $D_2 \in f(C_2)$ такі, що результат підстановки деякої m -резольвенти D_1 і D_2 належить D_3 ; 2) $f(\emptyset) = \{\emptyset\}$; 3) якщо результат деякої підстановки m -клаузи C_1 належить m -клаузі C_2 , то для будь-якої абстракції D_2 для C_2 є абстракція D_1 для C_1 така, що результат підстановки D_1 належить D_2 . Назвемо таке відображення f m -абстрактним відображенням, а будь-яке D із $f(C)$ – m -абстракцією. Під відображенням типізації будемо розуміти деяке відображення ϕ із літералів в літерали, яке відображає кожну атомарну формулу в формулу, терми якої мають тип, найближчий за ієрархією до базових типів. У праці [3] встановлено, що відображення типізації ϕ є m -відображенням.

Упорядковані лінійні виводи Tm і Um знаходяться у відношенні \rightarrow_f (позначимо $Tm \rightarrow_f Um$), де f – m -відображення, якщо вершини виводів Tm і Um знаходяться в такому відношенні відповідності R , що: 1) одночасно виконується $\forall n (n \in s - N(Tm)) \exists n' (n' \in s - N(Um)) (nRn')$ і

$\forall n (n \in s - N(Um)) \exists n' (n' \in s - N(Tm)) \quad (nRn') ;$ 2) якщо nRn' , де $n \in s - N(Tm)$, $n' \in s - N(Um)$, то n і n' можуть бути початковими або термінальними вершинами тільки одночасно; 3) для термінальних вершин Tm і Um відношення R є відношенням один до одного; 4) якщо $\langle n_1, n_2, n_3 \rangle \in s-M(Tm)$, $\langle n'_1, n'_2, n'_3 \rangle \in s-M(Um)$ і $n_3Rn'_3$, то $n_1Rn'_1$ і $n_2Rn'_2$; 5) якщо n і n' початкові вершини виводів відповідно Tm і Um і nRn' , то $s-L(n') \in f(s-L(n))$; 6) якщо n і n' непочаткові вершини виводів Tm і Um , відповідно, і nRn' , то приклад $s-L(n')$ належить $f(s-L(n))$.

Отже, упорядковані лінійні виводи мають властивість мінімальності, тобто точно одну термінальну вершину і, якщо трійка $\langle n_1, n_2, n_3 \rangle$ належить мінімальному упорядкованому лінійному виводу, то жодна інша трійка не може містити вершину n_3 як третю компоненту, за виключенням трійки $\langle n_1, n_2, n_3 \rangle$.

Використаємо ці властивості упорядкованих лінійних виводів для побудови механізмів виведення, які реалізуються агентами. Після того, як користувач сформулював задачу, система буде вивід – по суті, перевіряючи можливість задоволення запиту за рахунок наявних ресурсів. Це завдання виконують Агенти-Оркестратори як важлива складова інтелектуального ядра системи. Маючи доступ до переліку всіх сервісів, їх описів і аксіом, які й специфікують можливості застосування сервісів, вони запускають процес формування виводу. Тоді й відбувається побудова ланцюжка сервісів, здатного забезпечити потрібний користувачу результат, розпочинаючи з отримання даних із відповідних їх джерел.

Упорядкований лінійний m -резолютивний вивід подаватимемо парою множин вершин $V = \{ k_1, k_2, k_3 \dots k_n \}$ і трійок вершин $T \{ \langle k_1, k_2, k_3 \rangle, \langle k_3, k_4, k_5 \rangle \dots \langle k_{n-2}, k_{n-1}, k_n \rangle \}$, де k_i – вершини дерева, k_n – *термінальна* вершина, результат роботи алгоритму. Вивід формується, приймаючи за початкові вершини виводу постумову сформульованої проблеми та як бічну вершину – сумісну аксіому з бази знань. Аксіома є сумісною, якщо послідовність

атомарних формул, які відповідають даній вершині виводу, або будь-яка її частина, є постумовою аксіомою. Третю вершину цієї трійки отримуємо застосуванням аксіом до формули постумови з врахуванням правил виведення. Третя вершина першої трійки стає першою вершиною другої трійки, і процес повторюється доти, поки не буде досягнуто термінальної вершини – передумови, вказаної в проблемі. Якщо атомарну формулу, що відповідає третій вершині трійки, може бути спрощено застосуванням одного з правил виведення, вона спрощується в наступній трійці, яка матиме всього дві вершини – вершину зі спрощеною формулою і вершину зі спрощеною формулою. Пошук сумісних аксіом відбувається шляхом зіставлення конструкта оброблюваної в даний момент постумови з конструктами аксіом з онтології. Таким чином, з онтології вибирається множина аксіом, описані якими сервіси оброблюють необхідний нам тип і формат об'єктів. Підбір і перебір аксіом під час роботи механізму виведення відбувається лише на даній множині. Якщо під час пошуку аксіом з бази знань для чергової вершини сумісних аксіом знаходиться декілька, то кожна з цих аксіом використовується для подальшої побудови своєї «паралельної» версії виводу. Таким чином, під час роботи механізму виведення може формуватися деяка кількість виводів різної довжини і складності, залежно від кількості аксіом в базі знань і комбінацій їх застосування до кожної вершини.

Після завершення роботи механізму виведення, з множини отриманих виводів вибирається вивід з найменшою кількістю трійок вершин і найменшою кількістю застосованих аксіом. Довші, циклічні та тупикові виводи відкидаються. Множина формул і клауз, які відповідають вершинам трійок виводу і відображають застосування аксіом і правил при їх побудові, є виходом механізму виведення.

Алгоритм роботи механізму виведення відображено на рисунку 2.7.

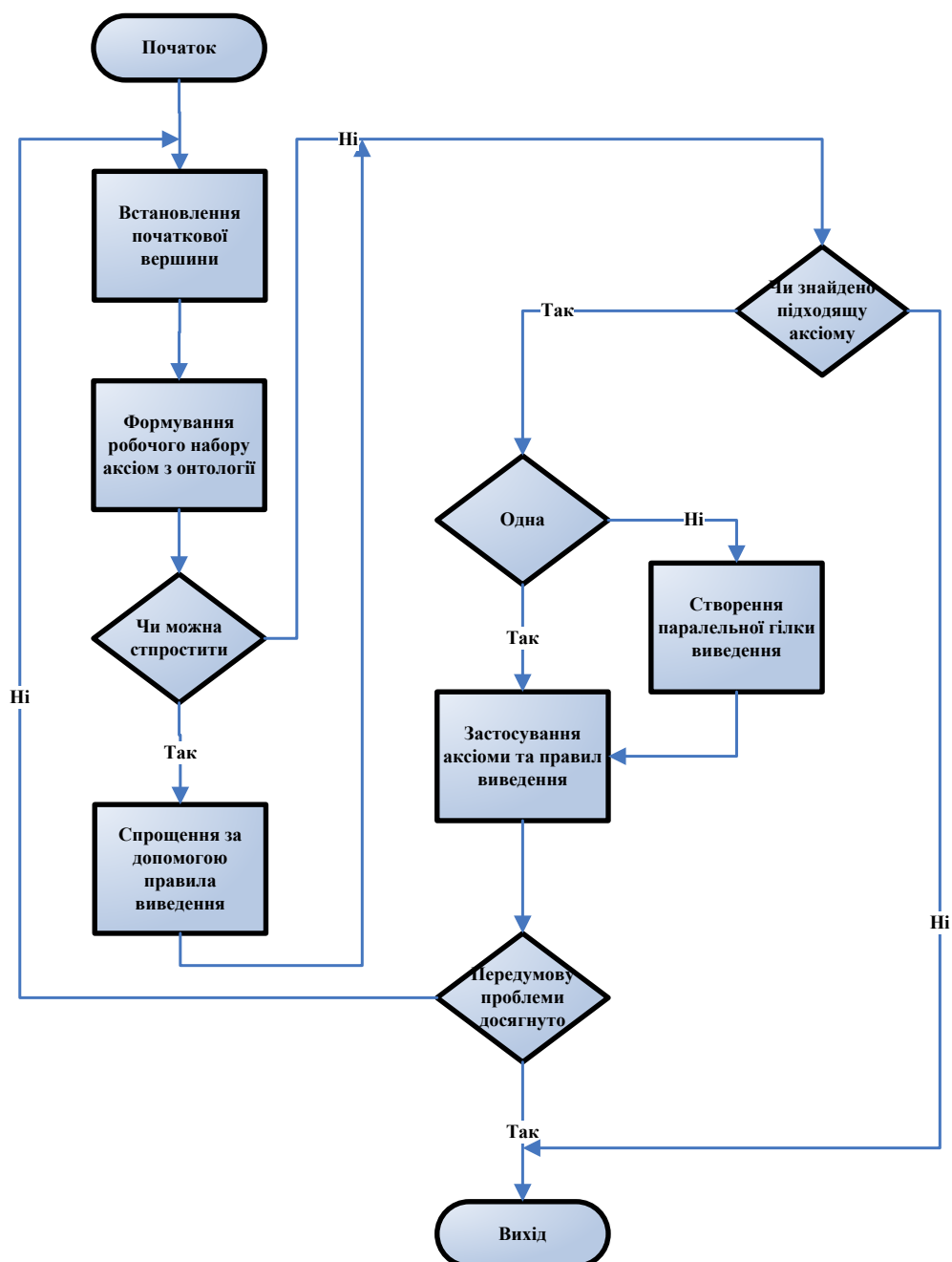


Рис. 2.7. Алгоритм роботи механізму виведення.

2.4. Механізм відновлення схеми вирішення проблеми на основі виводу

Для отримання з виходу механізму виведення функціональної послідовності дій (виконання сервісів або застосунків), яку і буде виконувати наша система, з врахуванням особливостей і характеру сервісів, необхідно задіяти механізм відновлення схеми рішення проблеми.

Схема вирішення становить зв'язний орієнтований граф без орієнтованих циклів з паралельними орієнтованими шляхами з кореня у

вершини, який має три типи вершин, і задається трійкою $G = \langle V, E, \Theta \rangle$, де $V = V_1 \cup V_2 \cup V_3$, V_1 – множина вершин-методів, V_2 – множина вершин-передумов і вершин-постумов, V_3 – множина вершин даних, в яких відбувається об'єднання або розщеплення даних; E – множина ребер; Θ – підмножина декартового добутку $E \times V \times V$, яка визначає співвідношення ребер і пар вершин. Схема визначає послідовність і взаємозв'язок по даним дій (сервісів, застосунків, які будуть виконані виконавчими механізмами системи з оброблення даних для отримання потрібного користувачу результату).

Алгоритм відновлення схеми вирішення. Дано: вивід $\text{Result} = \langle V, T \rangle$.

Знайти: $G = \langle V, E, \Theta \rangle$.

Крок 1. Визначення термінальної трійки – такої трійки, яка містить вершину k , що не є ані першою, ані другою складовою жодної з трійок множини T .

Крок 2. У вибраній трійці початкова вершина є вершиною постумови, друга – вершиною методу, термінальна – вершиною передумови. Метод, що перетворює передумову в постумову, визначається за однозначною відповідністю формулам вершин постумови та передумови аксіоми методу. Вершини з'єднуються ребрами передумова-метод, метод-постумова. У випадку, якщо передумова аксіоми методу складається з об'єднання або перетину декількох елементів, між вершинами передумов і методу вставляється вершина даних, що групує передумови згідно з відповідним правилом виведення. Якщо в термінальній трійці лише дві вершини, було застосоване правило виведення для перетворення клауз вершин, і місце вершини методу в цій гілці древа займає вершина даних перетворення, що відповідає застосованому правилу виведення, з відповідними вершинами постумов.

Крок 3. Вершини, ребра та їх співвідношення заносяться до G .

Крок 4. Вилучивши оброблену трійку з множини трійок вершин T , виконуємо крок 2. Якщо множина пуста – відновлення дерева вирішення завершено.

Після завершення відновлення дерева вирішення, воно подається на вхід виконавчого механізму. Виконання конкретного методу, вказаного в дереві вирішення, представлене конструктором, що описує вхідні дані (сутності, зв'язки, відношення між ними), передумови методу, постумови методу і вихідні дані. Гілки дерева вирішення, що знаходяться після вершини розщеплювання даних, можуть виконуватися паралельно доти, доки не дійдуть до вершини об'єднання даних, де, після завершення всіх паралельних гілок, залучених до об'єднання, продовжують виконуватися в послідовному режимі. Алгоритм роботи механізму відновлення дерева вирішення проблеми відображено на рисунку 2.8.

2.5. Приклад застосування логічного підходу до вирішення задач сталого розвитку регіонів України у СІД-Україна

Продемонструємо працездатність запропонованої в дисертаційній роботі реалізації логічного підходу на реальній науковій задачі розрахунку компоненти безпеки життя та індикації критичних значень показників загроз для аналізу сталого розвитку регіонів України [13].

Компоненти безпеки життя будемо обчислювати за формулою:

$$Csl = \sqrt[3]{\sum_0^n Threat^3}, \quad (2.1)$$

де значення загрози нормується за формулами (2.3) або (2.4).

Мета індикації критичних значень показників загроз полягає в визначенні їх пріоритетів в сенсі першочергової уваги, яку необхідно приділяти ним при прийнятті управлінських рішень на рівні окремого регіону та країни в цілому з метою зменшення впливу загроз на сталий розвиток.

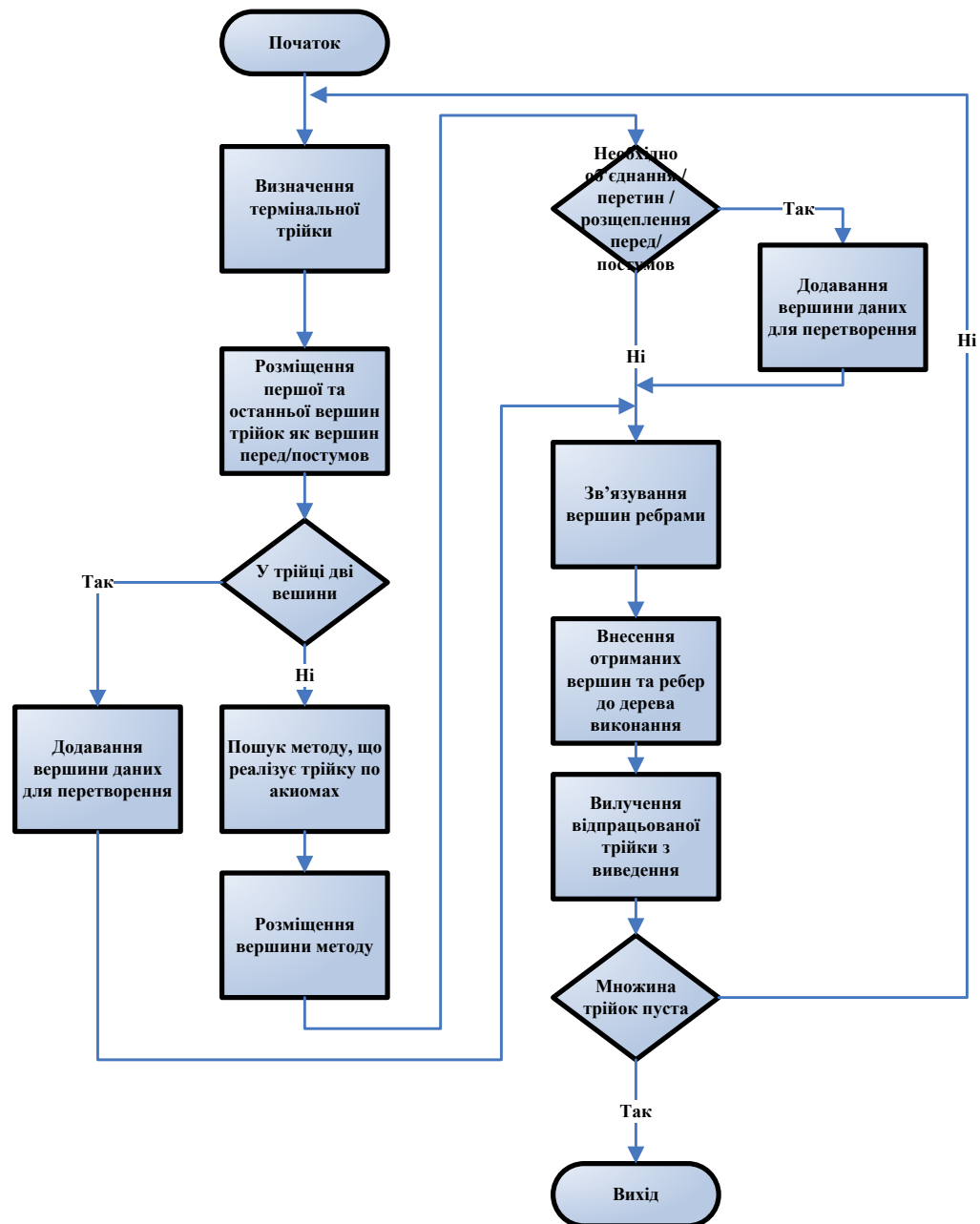


Рис. 2.8. Алгоритм роботи механізму відновлення вирішення проблеми

Нехай для кожної адміністративної одиниці $i = \overline{1, n}$ відомий набір значень $\langle x_{i,1}, x_{i,2}, \dots, x_{i,m} \rangle$ показників $X_j, j = \overline{1, m}$, які характеризують негативний вплив деяких явищ в економічній, соціальній та екологічній сфері на процеси сталого розвитку. Такі показники, сутність та склад яких визначається експертним шляхом, будемо називати показниками загроз.

Виходячи із змісту задачі індикації критичних значень, необхідно визначити таку характеристичну функцію:

$$\Psi(x_{i,j}) = \begin{cases} 0, & \text{якщо значення } x \text{ не є критичним;} \\ 1, & \text{навіпаки.} \end{cases} \quad (2.2)$$

де $i = \overline{1, n}$, $j = \overline{1, m}$.

Зрозуміло, що визначення функції $\Psi(x_{i,j})$ повинно спиратися на певні критерії, які враховують перевищення $x_{i,j}$ небезпечного порогу, відносно положення регіону i в рейтингу за показниками X_j , який складено для групи порівняння та для країни в цілому, ступінь «небезпечності» значення $x_{i,j}$ по відношенню до значень інших показників для регіону i .

Для врахування відносного положення регіону в рейтингах для країни в цілому використовується такий критерій:

$$R_{i,j} = \left(1 + e^{\frac{a-x_{i,j}}{b}} \right)^{-1}, \quad (2.3)$$

у випадку, коли більші значення показника X_j відповідають більшому впливу відповідної загрози на сталий розвиток, та:

$$R_{i,j} = 1 - \left(1 + e^{\frac{a-x_{i,j}}{b}} \right)^{-1} \quad (2.4)$$

у випадку, коли менші значення показника X_j відповідають більшому впливу.

У формулах (2.3, 2.4) параметри a і b обчислюються за формулами

$$a = \overline{X_j} = \frac{1}{n} \sum_{i=1}^n x_{i,j}, \quad b = \sigma(X_j) = \sqrt{\frac{\sum_{i=1}^n (x_{i,j} - \overline{X_j})^2}{n}}. \quad (2.5)$$

Критерій $R_{i,j}$ є безрозмірною величиною та приймає значення в діапазоні $[0,1]$. При цьому значення $R_{i,j}$ близькі до 0,5 відповідають середнім значенням X_j у вибірці, а значення більші від 0,75 відповідають значенням, які перевищують середні більше ніж на стандартне відхилення. У цьому випадку характеристична функція (2.2) з урахуванням одного критерію $R_{i,j}$ може бути записана у вигляді

$$\Psi_R(x_{i,j}) = \begin{cases} 0, R_{i,j} < 0,75; \\ 1, R_{i,j} \geq 0,75. \end{cases}$$

Обчислення критерію $P_{i,j}$, який враховує відносне положення регіону в групі порівняння, може бути здійснений згідно формули (2.3, 2.4) з урахуванням того, що параметри a і b обчислюються по формулі (2.5) для кожної групи порівняння окремо.

Критерії $R_{i,j}$ і $P_{i,j}$ є безрозмірними величинами та мають подібну природу, тому можуть бути агреговані за допомогою зваженої суми:

$$K_{i,j} = w_R R_{i,j} + w_P P_{i,j}; w_R + w_P = 1. \quad (2.6)$$

У формулі (2.6) вагові коефіцієнти w_R та w_P визначаються експертним шляхом. Таким чином, для регіону $i = \overline{1, n}$ маємо набір $\langle K_{i,1}, K_{i,2}, \dots, K_{i,m} \rangle$ значень агрегованого критерію, який враховує відносне положення регіону в рейтингах, складених для груп порівняння та для країни в цілому. Тепер серед значень $K_{i,j}$, $j = \overline{1, m}$ треба визначити найгірші значення, що можна зробити також за допомогою формули (2.3) що приводить до формули:

$$I_{i,j} = \left(1 + e^{\frac{a - K_{i,j}}{b}} \right)^{-1}, \quad (2.7)$$

У формулі (2.7) параметри a і b обчислюються на підставі значень вибірки $K_{i,j}$, $j = \overline{1, m}$. Оскільки для значень критерію $I_{i,j}$ справедливі всі зауваження, висунуті до критерію $R_{i,j}$, характеристична функція (2.2) може бути записана у вигляді

$$\Psi_I(x_{i,j}) = \begin{cases} 0, I_{i,j} < 0,75; \\ 1, I_{i,j} \geq 0,75. \end{cases}$$

Таким чином, значення характеристичної функції $\Psi_I(x_{i,j}) = 1$ відповідають найвищому пріоритету уваги, яка повинна приділятися

значенню показника X_j при прийнятті управлінських рішень на рівні окремого регіону i .

Покажемо як, використовуючи запропоновані моделі і методи математичної логіки, можна будувати дерево, яке визначає сервіси (застосунки), які треба викликати, порядок виклику і параметри, які треба передавати сервісам (застосункам) і отримувати від них для збереження, передавання чи опрацювання.

Завдання: розрахувати індикацію критичних значень показників загроз для країни в цілому, групи порівняння макрорайону та по відношенню до окремого регіону.

Дано: $X_{i,j}$ – набір значень показників загроз.

Проблема: $\langle\langle X_{i,j} \rangle, \langle R_{i,j}, R'_{i,j}, P_{i,j}, P'_{i,j}, I_{i,j}, I'_{i,j} \rangle\rangle$ де $R'_{i,j}, P'_{i,j}, I'_{i,j}$ критерії, які пройшли індикацію.

Аксіоми бази знань СЦД (ССД), які описують відповідні сервіси (застосунки), розгорнуті в СЦД (різних СЦД ССД):

1. Сервіс (застосунок) лінійного нормування для країни в цілому: задається аксіомою 1 ($\langle D_{i,j} \rangle, \langle R_{i,j} \rangle$).

2. Сервіс (застосунок) лінійного нормування для макрорайону: задається аксіомою 2 ($\langle D_{i,j} \rangle, \langle P_{i,j} \rangle$).

3. Сервіс (застосунок) агрегування критеріїв $R_{i,j}, P_{i,j}$: задається аксіомою 3 ($\langle R_{i,j}, P_{i,j} \rangle, \langle K_{i,j} \rangle$).

4. Сервіс (застосунок) лінійного нормування для окремого регіону: задається аксіомою 4 ($\langle K_{i,j} \rangle, \langle I_{i,j} \rangle$).

5. Сервіс (застосунок) індикації критичних значень загроз для країни: задається аксіомою 5 ($\langle R_{i,j} \rangle, \langle R'_{i,j} \rangle$).

6. Сервіс (застосунок) індикації критичних значень загроз для макрорайону: задається аксіомою 6 ($\langle P_{i,j} \rangle, \langle P'_{i,j} \rangle$).

7. Сервіс (застосунок) індикації критичних значень загроз для регіону:
задається аксіомою 7 ($\langle I_{i,j} \rangle, \langle I'_{i,j} \rangle$).

Пошук рішення із застосуванням аксіом бази даних і правил виведення за допомогою наведеного у підрозділі 2.4 методу виведення в клаузальній логіці взаємодії застосунків (вивід графічно представлений на рис. 2.9):

Аксіома 7: ($\langle I_{i,j} \rangle, \langle I'_{i,j} \rangle$)

Спрощення за допомогою правила 5:

$$R_{i,j}, R'_{i,j}, P_{i,j}, P'_{i,j}, I_{i,j}, I_{i,j} \Leftrightarrow R_{i,j}, R'_{i,j}, P_{i,j}, P'_{i,j}, I_{i,j}$$

Аксіома 4: ($\langle K_{i,j} \rangle, \langle I_{i,j} \rangle$)

Аксіома 3: ($\langle R_{i,j}, P_{i,j} \rangle, \langle K_{i,j} \rangle$)

Спрощення за допомогою правила 5:

$$R_{i,j}, R'_{i,j}, P_{i,j}, P'_{i,j}, R_{i,j}, P_{i,j} \Leftrightarrow R_{i,j}, R'_{i,j}, P_{i,j}, P'_{i,j}, P_{i,j}$$

Спрощення за допомогою правила 5: $R_{i,j}, R'_{i,j}, P_{i,j}, P'_{i,j}, P_{i,j} \Leftrightarrow R_{i,j}, R'_{i,j}, P_{i,j}, P'_{i,j}$

Аксіома 5: ($\langle R_{i,j} \rangle, \langle R'_{i,j} \rangle$)

Спрощення за допомогою правила 5: $R_{i,j}, R_{i,j}, P_{i,j}, P'_{i,j} \Leftrightarrow R_{i,j}, P_{i,j}, P'_{i,j}$

Аксіома 1: ($\langle D_{i,j} \rangle, \langle R_{i,j} \rangle$)

Аксіома 6: ($\langle P_{i,j} \rangle, \langle P'_{i,j} \rangle$)

Спрощення по правилу 5: $D_{i,j}, P_{i,j}, P_{i,j} \Leftrightarrow D_{i,j}, P_{i,j}$

Аксіома 2: ($\langle D_{i,j} \rangle, \langle P_{i,j} \rangle$)

Спрощення по правилу 5: $D_{i,j}, D_{i,j} \Leftrightarrow D_{i,j}$

Рішення знайдено. Тепер запишемо отриманий упорядкований лінійний вивід у вигляді множини відповідних трійок $\langle n_1, n_2, n_3 \rangle$ вершин:

$\langle D; D, D \rangle, \langle D, D; (\langle D \rangle, \langle P \rangle); D, P \rangle, \langle D, P; D, P, P \rangle, \langle D, P, P; (\langle P \rangle, \langle P' \rangle); D, P, P' \rangle, \langle D, P, P'; (\langle D \rangle, \langle R \rangle); R, P, P' \rangle, \langle R, P, P'; R, R, P, P' \rangle, \langle R, R, P, P'; (\langle R \rangle, \langle R' \rangle); R, R', P, P' \rangle, \langle R, R', P, P', R, R', P, P', P \rangle, \langle R, R', P, P', P; R, R', P, P', R, P \rangle, \langle R, R', P, P', R, P; (\langle R, P \rangle, \langle K \rangle); R, R', P, P', K \rangle, \langle R, R', P, P', K; (\langle K \rangle, \langle I \rangle); R, R', P, P', I \rangle, \langle R, R', P, P', I; R, R', P, P', I, I \rangle, \langle R, R', P, P', I, I; (\langle I \rangle, \langle I' \rangle); R, R', P, P', I, I' \rangle$

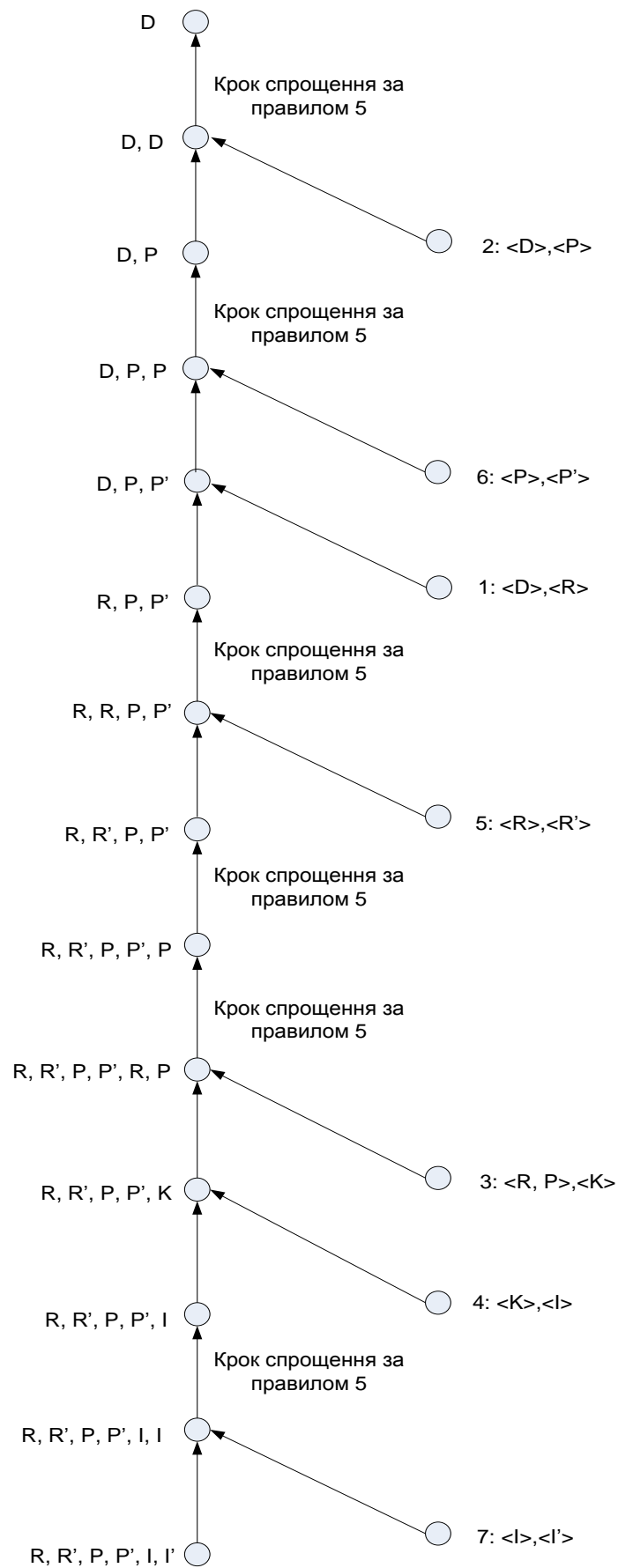
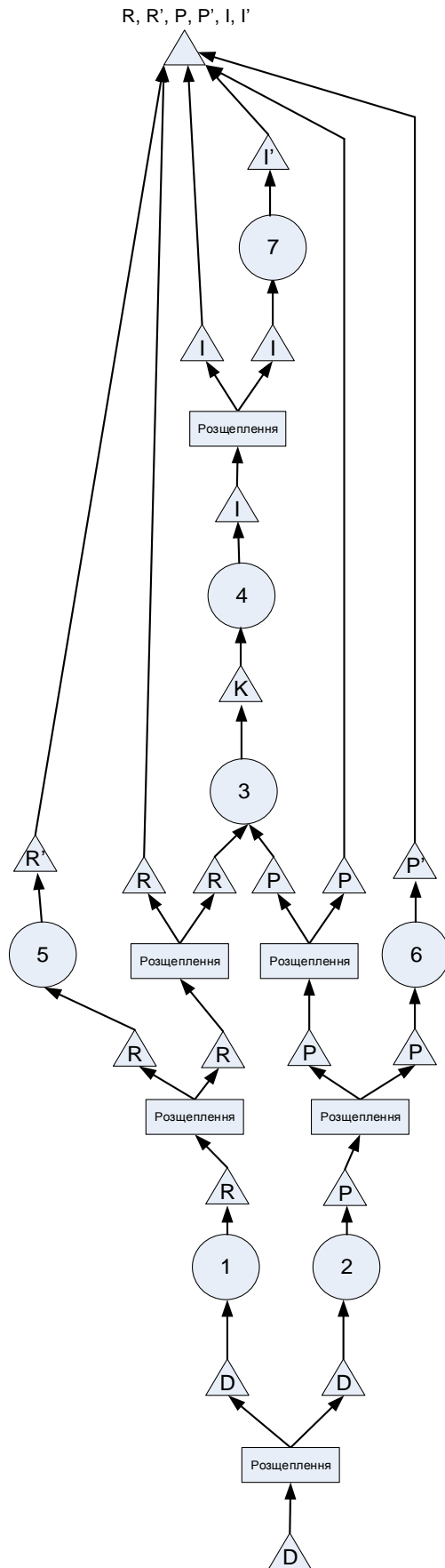


Рис. 2.9. Вивід для проблеми $\langle\langle X_{i,j} \rangle, \langle R_{i,j}, R'_{i,j}, P_{i,j}, P'_{i,j}, I_{i,j}, I'_{i,j} \rangle\rangle$



Тепер для того щоб визначити порядок виконання сервісів (застосунків), їх виклику і передачі відповідних параметрів відновимо дерево вирішення проблеми за цими трійками отриманого виводу, застосовуючи механізм відновлення дерева вирішення проблеми, описаний у підрозділі 2.4. Дерево вирішення проблеми наведене на рисунку 2.10.

Знайдена послідовність виконання сервісів (застосунків), які визначаються аксіомами бази знань, що їх описують:

1. Аксіома 2: ($\langle D_{i,j} \rangle, \langle P_{i,j} \rangle$)
2. Аксіома 6: ($\langle P_{i,j} \rangle, \langle P'_{i,j} \rangle$)
3. Аксіома 1: ($\langle D_{i,j} \rangle, \langle R_{i,j} \rangle$)
4. Аксіома 5: ($\langle R_{i,j} \rangle, \langle R'_{i,j} \rangle$)
5. Аксіома 3: ($\langle R_{i,j}, P_{i,j} \rangle, \langle K_{i,j} \rangle$)
6. Аксіома 4: ($\langle K_{i,j} \rangle, \langle I_{i,j} \rangle$)
7. Аксіома 7: ($\langle I_{i,j} \rangle, \langle I'_{i,j} \rangle$)

Рис. 2.10. Відновлене дерево вирішення проблеми $\langle\langle X_{i,j} \rangle, \langle R_{i,j}, R'_{i,j}, P_{i,j}, P'_{i,j}, I_{i,j}, I'_{i,j} \rangle\rangle$

Висновки за розділом

1. Запропонована цілісна концепція інтеграції застосунків і джерел даних в рамках платформи підтримки міждисциплінарних досліджень у Світовій системі даних.

2. Розроблено комплексний підхід до виконання сформульованого користувачем запиту. Ключові рішення підходу: для створення реєстру сервісів використовується UDDI; для уніфікованого описання сервісів використовується WSDL; обмін повідомленнями реалізується за допомогою SOAP; оркестрація сервісів здійснюється через BPEL; для описання агентів використовується JADE.

3. Розроблено архітектуру MAC для перетворення неоднорідної, але семантично тотожної інформації, в метадані, придатні для наступного аналізу і дослідження.

4. Розроблена формальна логічна система – формалізм, на базі якого буде побудована інтеграційна складова платформи. Описані складові формальної системи: мова клаузальної логіки першого порядку; база знань, основу якої складають онтологія аксіом, які зображують методи сервісів системи, і онтології джерел даних, описаних на мові OWL на базі RDF; правила виведення, необхідні для отримання потрібного результату у випадках, коли необхідно комбінувати методи як одного, так і різних сервісів.

5. Запропоновано метод виведення, побудований на типізації тверджень і аналогії, як формальну основу механізму виведення у складі платформи. Побудований в абстрактному просторі вивід використовується для управління процесом виведення у вихідному просторі пошуку розв'язків, що дозволить підвищити ефективність виведення за рахунок відсікання більшої частини безперспективних гілок виводу у вихідному просторі.

6. Розроблено алгоритм відновлення схеми вирішення проблеми для інтеграції сервісів (застосунків) як основу механізму відновлення схеми вирішення для отримання з виходу механізму виведення функціональної послідовності дій, яку і буде виконувати система, з врахуванням особливостей

і характеру сервісів (застосунків).

7. Наведено приклад застосування логічного підходу до вирішення задач сталого розвитку регіонів України у СЦД-Україна. Продемонстрована працездатність запропонованої в дисертаційній роботі реалізації логічного підходу на реальній науковій задачі розрахунку компоненти безпеки життя та індикації критичних значень показників загроз для аналізу сталого розвитку регіонів України.

РОЗДІЛ 3.

МОДЕЛІ І МЕТОДИ ІНТЕГРАЦІЇ ДЖЕРЕЛ ДАНИХ І СИСТЕМНОГО УЗГОДЖЕННЯ ДАНИХ РІЗНОЇ ПРИРОДИ

3.1. Моделі і методи інтеграції джерел даних в рамках платформи підтримки міждисциплінарних досліджень в Світовій системі даних

Описаний у розділі 2 підхід до вирішення проблеми інтеграції джерел даних визначає структуру і принципи функціонування системи інтеграції багатьох джерел даних при розв'язанні міждисциплінарних наукових задач. Структура і принципи функціонування, у свою чергу, обумовлюють вибір математичних моделей і методів, насамперед через вимоги і обмеження. Використання дескриптивних логік для описання джерел даних, модульність, готовність до приєднання ланцюжків сервісів для додаткового оброблення даних приводять до розподіленої системи з центральним вузлом, який працює на сервері застосунків як Java Enterprise компонент с налаштованою точкою доступу. Для приєднання кожного джерела даних добавляються OWL-файл з семантичною маскою метаданих і файл властивостей для безпосереднього підключення зовнішнього джерела з уточненням типу драйвера, IP-адреси тощо. Зрозумілою також стає роль користувацького інтерфейсу і системи конструювання запиту.

Склад компонентів і послідовність їх роботи над запитом наведені на рисунку 3.1. Компоненти системи функціонують на основі моделей, розроблених при виконанні дисертаційного дослідження і описаних нижче.

Спочатку розглянемо етапи виконання запитів як цілісний процес, а потім детальніше розглянемо кожний з етапів з метою визначення вимог і обмежень, важливих для побудови нових, модифікації чи розвитку існуючих моделей та методів, які будуть покладені в основу реалізації відповідних компонентів системи інтеграції джерел даних платформи підтримки міждисциплінарних досліджень.

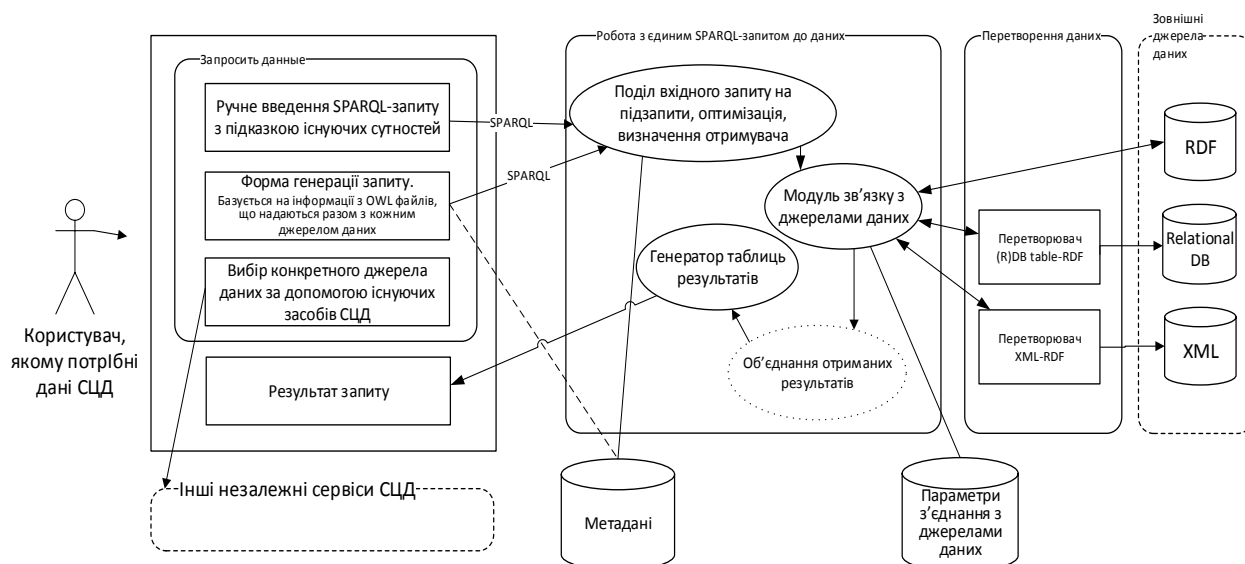


Рис. 3.1. Склад і послідовність роботи компонентів системи інтеграції джерел даних

Етап 1. Генерування SPARQL-запитів за допомогою конструктора або отримання готового запиту від користувача.

Етап 2. Парсинг, переписування і оптимізація запиту, які виконуються на основі онтологічного опису джерел даних та метаданих.

Етап 3. Розбиття запиту на підзапити, які спрямовуються безпосередньо до джерел даних.

Етап 4. Виконання підзапитів на вибраних джерелах даних. Запити оброблюються адаптером доступу до джерел даних, які функціонують автономно і отримавши запит, виконують його і повертають результат у вигляді набору записів.

Етап 5. Очікування отримання відповідей від усіх джерел даних.

Етап 6. Об'єднання отриманих результатів за допомогою Inner Join (опціонально). Над отриманими від джерел даних результатами виконання запитів, виконуються необхідні операції згідно сформованого плану виконання запиту. Застосовуються стандартизовані адаптери, специфічні адаптери, формат онтологій з метаданими, описаними у XML-подібному синтаксисі. Це дозволяє охопити широке коло джерел.

Етап 7. Повернення користувачу єдиної таблиці або множини таблиць

(якщо об'єднання неможливе або користувач вирішив його не виконувати) як результат виконання запиту.

Тепер детальніше розглянемо моделі і методи, запропоновані в роботі для реалізації компонентів платформи, що підтримують наведені етапи.

Розпочнемо із моделей описання джерел даних на семантичному рівні представлення.

Інтеграція неоднорідних інформаційних ресурсів, вимоги до реалізації якої сформульовані у підрозділі 2.1, вимагає описання структури і поведінки кожного інформаційного ресурсу. Структура і поведінка ресурсів визначаються семантикою відповідних предметних областей. Виходячи з концепції інтеграції джерел даних, семантику описуватимемо на основі дескриптивної логіки, базовими елементами якої є:

- множина класів NC;
- множина індивідуумів NI;
- множина відношень NR.

Як показано у дисертації, на цій основі можна описати класи, ієрархії, відношення, властивості, типи й екземпляри, ресурси. Для визначення функцій і аксіом будемо використовувати інші мови.

Насамперед постає потреба в зв'язній структурі даних із онтологій та підключень до джерел даних. Тобто необхідно побудувати у пам'яті застосунків зв'язну ієрархію із онтологій-класифікатора, розширюючих онтологій, онтологій джерел даних, онтологій, що зв'язують C-властивості джерел, а також їх співставлення з описом відображення джерел даних.

У роботі розроблено цілісний підхід до побудови зв'язної структури даних із онтологій та підключень до джерел даних, який включає такі етапи.

Етап 1. Завантаження онтологій-класифікатора, побудова відповідного головного дерева C-класів.

Етап 2. Послідовне завантаження розширюючих онтологій з під'єднанням їх у пам'яті до головного дерева. Послідовність завантаження розширюючих онтологій може бути визначена чергою (це необхідно,

наприклад, для випадку, коли одна розширююча онтологія розширює іншу).

Етап 3. Послідовне завантаження онтологій-описів джерел даних. При цьому такі онтології можуть мати декілька рівнів підкласів, що робить їх порівнюваними з розширюючими онтологіями.

Етап 4. Приєднання завантажених онтологій-описів джерел даних до головного дерева, співставлення співпадаючих С-класів і їх С-властивостей та визначення їх (за наявності) як таких, що доступні у декількох джерелах даних.

Етап 5. Завантаження зв'язуючої онтології, проекція її на збудоване головне дерево С-класів і С-властивостей. Результатом є граф, який пов'язує певні листки (С-властивості) збудованого головного дерева.

Етап 6. Завантаження описів відображень джерел даних і встановлення їх зв'язків з С-класами та С-властивостями.

Приклад побудованого в пам'яті застосунку дерева С-класів, С-властивостей, їх зв'язків між собою та з джерелами даних наведений на рис. 3.2.

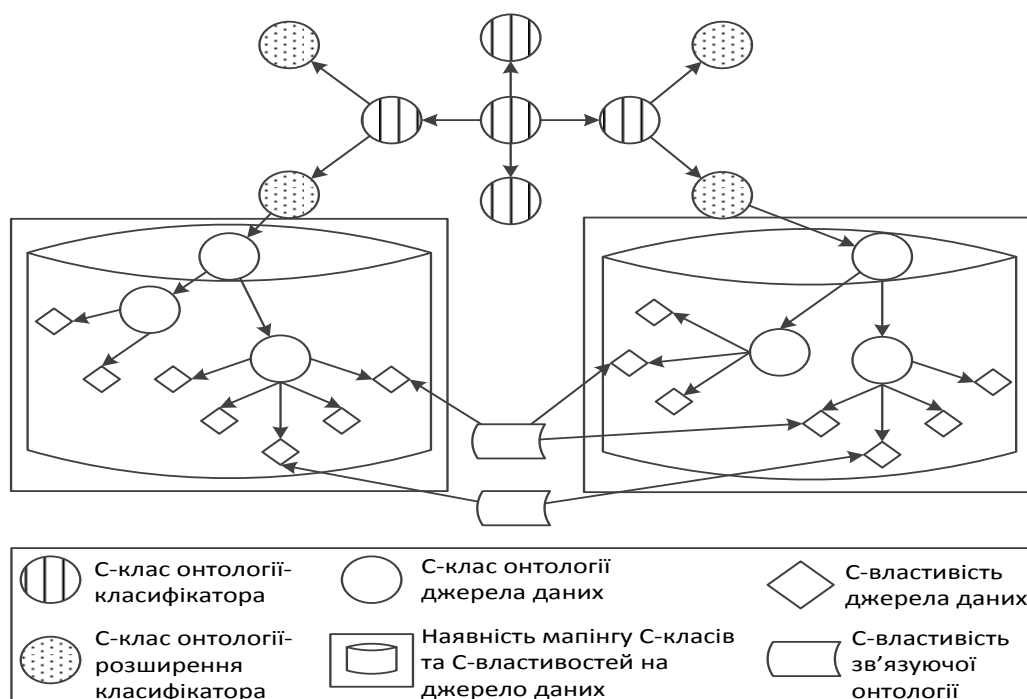


Рис. 3.2. Приклад об'єктної моделі об'єднаних онтологій та описів джерел даних

Тепер перейдемо до моделей підготовки описів джерел даних і їх перетворення у формат RDF. Розпочнемо з бібліотеки для організації мапінгу

не RDF-даних на онтологію. Для інтеграції семантики предметної області з джерелом даних, необхідно реалізувати механізм мапінгу концептів онтології відповідним об'єктам джерела даних.

Таку задачу здатна здійснити бібліотека Java D2R, яка забезпечує доступ до даних через SPARQL, Jena API і Sesame API у вигляді RDF-графу. Бібліотека D2R забезпечує зв'язок з джерелом даних через JDBC драйвер, що дозволяє використовувати в якості джерела практично будь-який ресурс – від СУБД до плоских файлів і HTML-документів, а при необхідності реалізувати специфічний драйвер. Роль носія семантики предметної області відіграє OWL-онтологія. Файл мапінгу будується в вигляді RDF-документа в форматі N3, доступ до якого може бути реалізовано методами Jena, а отже – автоматизовано його створення.

На основі такого опису платформа підтримки міждисциплінарних досліджень забезпечить користувачеві інтегрований доступ до метаданих, що описують семантику предметної області та джерела даних. При цьому спрощується взаємодія з клієнтською частиною системи. Зокрема, її реалізація в вигляді web-застосунку виключить необхідність встановлення додаткового програмного забезпечення та його налаштування.

Наступним важливим кроком є побудова моделей семантичного рівня для реалізації можливості роботи з описами джерел і інформаційних ресурсів, побудованих на доступі до джерел даних. Обґрунтовані вище і прийняті рішення однозначно визначають використання для опису джерел даних RDF-графів. Тепер необхідно вибрати інструменти для створення і маніпулювання RDF-графами. Таким засобом вибрано Jena [107], що є Java API застосунком. Jena має класи об'єктів для представлення графів, ресурсів, властивостей і літералів. Інтерфейси представлення ресурсів, властивості і літералів називаються відповідно Resource, Property та Literal. У Jena, граф називається моделлю і представлений інтерфейсом Model.

Таким чином, каркас Jena надає високорівневий доступ до графів RDF-документу і тим самим простий і гнучкий спосіб завантаження онтології,

маніпулювання триплетами, що в ній містяться, об'єднання декількох графів, наприклад онтології та RDF-документу.

Для отримання відповіді на запит користувача шляхом побудови схеми виклику сервісів (застосунків) і передачі їм відповідних параметрів у розділі 2 була запропонована клаузальна логіка і механізми виведення і відновлення вирішення. Другим підходом до формування відповіді на запити користувача на основі ефективного використання семантики предметної області є реалізація логічного виведення на базі побудованого графу. Ця схема доповнює можливості інтеграції застосунків і джерел, розширюючи спектр інструментів дослідників, як у отриманні потрібних даних, так і у швидкому розробленні і представленні доступу до інформаційних ресурсів.

Для розв'язання другого підходу було вибрано машину логічного виведення Pellet [108]. Pellet реалізує логічне виведення на RDF-даних з OWL DL (дескриптивна логіка). Ця система включає в себе оптимізацію для номіналів, відповіді на кон'юнктивні запити, а також послідовне виведення на нових графах.

Pellet підтримує логічне виведення і надає функції перевірки цілісності, когерентності класу, класифікації для відповіді на запити, такі як отримання всіх або тільки прямих підкласів класу, визначення прямого типу для індивідів. Важливо, що Pellet має API для спільної роботи з інструментом Jena.

Тепер перейдемо до моделей генерації єдиного запиту на семантизовані дані. Розглянемо генерацію первинного запиту на дані, сформульованого користувачем. Для цього можна використовувати такі підходи:

- 1) користувач самостійно вводить запит на мові запитів SPARQL;
- 2) запит конструюється користувачем на основі попередньо розробленої моделі запиту у просторі імен;
- 3) запит конструюється за допомогою візуальних засобів, головною функцією яких є надання інтерфейсу для побудови семантичних графів-трійок «суб'єкт-предикат-об'єкт» і обмежень на них;

4) користувач вводить запит природною мовою, а система надає засоби його розпізнавання і переведення в формат SPARQL;

5) пошук за ключовим словом і його розвиток – інкрементне конструювання запиту.

У роботі розроблено і реалізовано моделі для перших трьох підходів.

Розпочнемо з першого з них. Оскільки самостійне введення SPARQL-запиту до наявних джерел даних передбачає високий рівень компетентності користувача, в роботі розроблено модель для зручного для користувача вибору складових запиту і відповідний композитний алгоритм підтримки.

Введемо необхідні для реалізації підходу і описання алгоритму позначення. Нехай $A = \{A_1, \dots, A_m\}$ – множина доступних в системі сутностей в рамках глобального простору імен, $B = \{B_1, \dots, B_m\}$ – множина обмежень на діапазони даних в $A_i, i=1, \dots, m$, $B_i, i=1, \dots, m$, може бути порожнім, m – кількість доступних в системі сутностей в рамках глобального простору імен, $C \subset A$ – підмножина сутностей із множини A , яка визначається користувачем в процесі формування запиту і уточнює сутності, інформація про які має бути результатом пошуку.

Для уточнення операторів опрацювання вибірки даних введемо множини D, G, E, \dots параметрів цих операторів, наприклад LIMIT, ORDER BY, OFFSET і т.і.

Опишемо запропонований алгоритм підтримки побудови запиту:

Крок 1. Користувач вибирає сутності і обмеження на них. Запит на дані подається у вигляді кортежу:

$\langle A[G], B[G], C[G], \text{опціональні параметри} \rangle$

Отриманий опис реалізований у вигляді веб-форми з вибором відповідних даних, а наведена у цій формі інформація очевидним чином конвертується в SPARQL-запит, оскільки сутності кортежу є одночасно і сутностями цього запиту.

Крок 2. Оброблювач запитів отримує на вході запит, згенерований на кроці 1, і аналізує кількість джерел даних, в яких є інформація про потрібні

сутності. Якщо джерел даних декілька, то користувачу пропонується уточнити запит, вибравши відповідне джерело даних. У цьому випадку користувач отримує стислу інформацію щодо кожного із можливих джерел даних для цього запиту. Ця інформація витягується безпосередньо із джерела даних або із OWL-файлу зі стандартизованим форматом метаданих. Якщо користувач не може уточнити джерело даних, запит буде виконаний кожним можливим джерелом даних.

Для реалізації другого підходу розроблено модель формування єдиного запиту до множини джерел даних користувачем у єдиному просторі імен. Користувач оперує ієрархією С-класів, зв'язками між ними (описані owl-предикатами SubclassOf і SameAs), а також С-властивостями, які можуть бути у кожного С-класу, і обмеженнями на ці С-властивості. На рівні логіки застосунку відомо, які джерела даних пов'язані з якими С-класами і С-властивостями, але для користувача це не має ніякого значення.

Послідовність дій при формуванні запиту:

- 1) вибір необхідного С-класу в дереві класифікатора;
- 2) встановлення обмежень на значення його С-властивостей (за їх наявності) за допомогою констант або логічних виразів;
- 3) (опціонально) встановлення зв'язків з С-властивостями інших С-класів з формуванням Join-запиту.

Таким чином, на виході може бути отримано один або більше одного С-класів, набір С-властивостей і обмежень на них, С-властивості, через які треба об'єднувати значення для отримання кінцевого результату. У роботі реалізована підтримка чотирьох типів стандартних запитів до розподілених джерел даних, які можуть бути оброблені в межах запропонованої моделі:

- 1) *запит до однієї сутності* – часто вживаний запит на отримання підмножини набору даних з одного джерела даних за деякою умовою. Класичним прикладом такого запиту є Select-подібний запит. У випадку запиту до даних з інтегрованою семантикою такий SPARQL-запит буде містити URI потрібного С-класу, за яким закріплене тільки одне джерело

даних, і при цьому з інформації про діапазонах допустимих значень (при її наявності) впливає, що запит може бути виконаний на цьому джерелі;

2) *запит, в якому сутність рознесена в декількох джерелах даних (Union)* – запит на отримання за деякою умовою даних результату об'єднання С-властивостей С-класів (які можна представити як таблиці) згідно їх специфікації з різних джерел даних. Тоді йдеться про вхідний запит з одним С-класом і умовою. При цьому у декількох джерелах даних відображено зазначений С-клас і виконується частково або повністю зазначена умова;

3) *запит з об'єднанням сутностей (Join)* – запит на отримання перетину результатів кількох вибірок за С-властивостями різних С-класів, що може бути здійснено за наявності спільних С-властивостей у таких С-класів у зв'язуючій онтології. Зв'язки між зв'язуючими С-властивостями дозволяють об'єднати запити для утворення в результаті об'єднаного за допомогою Inner Join набору значень.

4) *запит на узагальнення* – запит на отримання з джерел даних наборів таблиць або описів С-класів, С-властивостей та їх зв'язків, які не підлягають об'єднанню, але, згідно опису в онтології-класифікаторі, вони є підкласами обраного користувачем С-класу. На практиці отримання повних таблиць даних з усіх підкласів обраного С-класу може бути недоцільним і створити надмірне навантаження на потужності серверів-провайдерів, і це питання може потребувати окремого вивчення.

Тепер перейдемо до моделі конструювання запиту за допомогою візуальних засобів, коли формуються запити безпосередньо до джерел даних. Поля таблиць джерела даних приводяться у абстрактну модель даних у вигляді графу «суб'єкт-предикат-об'єкт» і навпаки за допомогою опису відображень (мапінг). Представимо опис відображення у вигляді

$$f: R \leftrightarrow S,$$

де R – вихідна структура джерела даних, S – абстрактне семантичне представлення у вигляді графу.

Для конвертації запитів, виділених до конкретних джерел даних, в SQL-

формат з урахуванням властивостей і параметрів цих джерел згідно опису відображень використаємо алгоритм безпосередньої конвертації запитів, наведений у праці [109].

Насамкінець можемо зосередитися на *моделях поділу запиту на розподілені дані на підзапити, визначення плану виконання запиту і його оптимізації та виконання*. Саме ці моделі дозволяють побудувати ефективні алгоритми оброблення запитів до розподілених джерел даних, які, у свою чергу, визначають ефективність інтеграції джерел даних. За умови перетворення всіх імен даних у запиті у відповідності з глобальним простором імен для опрацювання єдиного запиту до розподілених джерел даних будемо використовувати описані в праці [38] рішення, які вже стали класичними.

Розроблені як частина проекту IBM Starburst вони закріплюють за оброблювачем, який отримав запит на вході три основних операції: аналіз і трансформацію запиту у план його виконання на основі наявної інформації щодо джерела (джерел) даних; багатокрокову оптимізацію запиту; виконання плану і отримання результатів запиту. Розглянемо моделі для кожної з цих операцій.

Запропонований підхід до виконання запитів до розподілених джерел семантизованих даних узагальнює напрацьовані раніше рішення. Для виконання запитів до розподілених RDF-джерел даних, які отримують на вході SPARQL-запити, найбільш прийнятна система інтеграції D2RQ [109] і моделі, наведені у праці [110]. Система інтеграції D2RQ пропонує єдиний інтерфейс для опитування множини розподілених SPARQL-кінцевих точок і здійснює федерування запиту прозоро для клієнта. Виконувана D2RQ оптимізація SPARQL-запитів, побудована на статистичній інформації щодо кількості триплетів у джерелі даних, також досить ефективна. Але D2RQ погано документований і не здатний самостійно (без участі адміністратора) завантажувати інформацію щодо типів даних (дескрипторів) джерела, які мають специфічний формат.

Тому в запропонованій системі інтеграції впроваджені моделі

федерування запитів до розподілених джерел даних, деякі алгоритми якої мають відмінності від розроблених для системи D2RQ:

- 1) застосування як інформації щодо типів даних для кожного з підключених джерел даних файлів OWL і N3, які також застосовуються для уточнення семантики предметної області і додавання метаданих;
- 2) забезпечення можливості виконання запитів не тільки на SPARQL-кінцевих точках, а й із застосуванням інших засобів комунікації, насамперед, обміну повідомленнями JMS або FIPA ACL;
- 3) функціонування в умовах відсутності можливості оптимізувати запит, коли відповідні статистичні дані з віддалених джерел не доступні або змінилися;
- 4) забезпечення можливості задавати цілий набір джерел даних, на які будуть направлятися згенеровані запити, що уможлиблюється запропонованим алгоритмом генерації запиту, коли користувач уточнив доступні для виконання запиту джерела даних.

Федерування запиту на семантизовані дані. Розпочнемо з описання даних й адаптуємо класичну модель поділу, запропоновану в D2RQ. Ємністю джерела даних D є множина C_D кортежів $c = (p, r) \in C_D$, де p – наявний в D предикат, r – обмеження на суб'єкти і об'єкти. Ця зв'язка є регулярним filter-виразом мови SPARQL, яка робить вибір джерела даних більш точним, наприклад, можна уточнити, що джерело даних зберігає інформацію тільки про специфічні типи ресурсів. Позначимо обмеження r як функцію $r(subject, object)$ з $r: (RDF_T \cup V) \times (RDF_T \cup V) \rightarrow \{true, false\}$.

Застосовувати такі обмеження можна, наприклад, у тому випадку, коли в одному джерелі даних зберігаються рядкові дані, які починаються з букв від А до R, тоді як в іншому розташовані дані, які починаються з букв від Q до Z.

Обмеження на шаблони доступу. Деякі джерела даних мають обмеження на шаблони доступу. Наприклад, джерело даних може вимагати, щоб в запиті завжди були наведені назва міста, в якому виконується пошук, або його поштовий індекс.

Застосуємо рішення, запропоноване в [109]. Оскільки предикати повинні бути обмеженими, вони використані як основа для складання шаблону. Предикат не може бути змінною. Нехай L_D становить множину обмежень на шаблони доступу для джерела даних D і $(S, O) \in L_D$ буде одним шаблоном з S і O , які повинні мати зв'язані суб'єкти (S) або зв'язані об'єкти (O).

Джерело D може відповідати на запит з графовим шаблоном P , якщо він задовольняє щонайменше один із позначених шаблонів доступу для D . Нехай $bound(x)$ буде функцією, яка повертає *false*, якщо x є змінною, і *true* в іншому випадку. Тоді шаблон доступу (S, O) задовольняється, якщо

$$(\forall p_s \in S \setminus O: \exists (s, p_s, o) \in P: bound(s)) \wedge (\forall p_o \in O \setminus S: \exists (s, p_o, o) \in P: bound(o)) \wedge (\forall p_b \in S \cap O: \exists (s, p_b, o) \in P: bound(s) \wedge bound(o))$$

Використання статистичної інформації. Наявність статистичної інформації щодо доступних даних допомагає оптимізатору запитів знайти план виконання запиту, ефективний за витратами. Але враховуючи незалежність і розподіленість джерел даних, збір статистичної інформації не завжди можливий. Ефективність системи інтеграції даних визначається компонентом планування виконання запиту та його оптимізації. Оскільки джерела даних функціонують автономно, статистичні дані про їх наповнюваність отримати не завжди можливо. До того ж доцільність виконання надскладних запитів до підключених джерел даних, враховуючи їх специфіку у ССД і невелику кількість поєднаних властивостей, є сумнівною.

Дієвим рішенням проблеми планування та оптимізації запитів є підхід на основі статистичної інформації, яку надають самі постачальники даних. Для неї виділено окреме поле метаданих в онтологічному описі джерела даних, яке заповнюється цілим числом, що вказує на орієнтовну кількість кортежів у джерелі даних в межах описуваної таблиці або перетину множини таблиць.

Таким чином, за необхідності побудови запиту двох і більше джерел даних, обробник генерує дерево-план виконання запиту, яке відпрацьовує запити до джерел з найменшою кількістю записів (рядків), передаючи в запиті до кожного наступного (пов'язаного з ним) джерела даних їх результат як

параметр і одержуючи їх об'єднання на виході. За умови виконання запитів до більше, ніж одного джерела даних – вони виконуються паралельно. Тоді на кожному кроці відкидається певна кількість нерелевантних записів, як при SQL-запиті Inner Join, а на виході – отримуємо єдину таблицю із результатом виконання запиту.

Для побудови раціонального плану виконання запиту із врахуванням коефіцієнтів ємності джерел даних у роботі розроблено відповідний алгоритм:

Крок 1. Отримання запиту у єдиному просторі імен, що визначає множину джерел даних і набір С-властивостей, за якими можуть бути об'єднані результати запитів до даних джерел даних, коефіцієнтів ємності кожного джерела даних.

Крок 2. Знаходження добутку коефіцієнтів ємності кожної пари джерел даних, підзапити до яких підлягають об'єднанню.

Крок 3. Виділення пари джерел даних, підзапити до яких підлягають об'єднанню.

Крок 4. Впорядкування цих пар у порядку зростання добутку коефіцієнтів ємності.

Крок 5. Виділення джерел даних, запити до яких можуть бути виконані паралельно (не потребують вхідних параметрів – результатів виконання запитів з інших джерел) та мають найменші коефіцієнти ємності.

Крок 6. Паралельне виконання запитів до виділених на кроці 5 джерел даних.

Крок 7. Виділення з отриманих на кроці 6 результатів даних, за якими виконується перетин із джерелом-парою та передавання як параметра в запиті до цього джерела-пари.

Крок 8. Представлення отриманого перетину результатів запитів до кожної пари джерел даних як віртуального джерела даних, коефіцієнт ємності якого дорівнює кількості кортежів отриманого результату, а запит вже було виконано.

Крок 9. Виконання кроків 3 – 8 доки не буде отримано єдиний кортеж,

який і стане результатом виконання запиту користувача.

Наповнення каталогу інформацією. На центральному вузлі системи (оброблювачі вхідних запитів) є каталог, який зберігає інформацію щодо всіх підключених сервісів.

Первинна інформація щодо даних у кожному джерелі вибирається із його опису – відповідних файлів N3 і OWL, які також використовуються для додавання семантичного забарвлення даним.

Планування виконання запиту. При формуванні запиту до розподілених джерел даних необхідно вирішити, яке із джерел здатне на цей запит відповісти. Процеси знаходження релевантних джерел і генерування виконуваних підзапитів і складають сутність планування запитів. Планування запитів побудоване на інформації, отриманій із Сервіс-дескриптора. Нехай $R = \{(d_1, C_1), \dots, (d_n, C_n)\}$, тут $d_1..d_n$ – множина джерел даних і $C_1..C_n$ будуть їх вмістом, де $C_i = \{(p_{i,1}, r_{i,1})..(p_{i,m}, r_{i,m})\}$.

Також на вибір доступних джерел даних впливає можлива наявність специфікованих користувачем додаткових обмежень при інкрементному генеруванні запиту – на яких саме джерелах даних необхідно виконувати запит.

Вибір джерела даних. Запит на мові SPARQL містить не менше одного з фільтрованих базових графових шаблонів (FBGP), кожний із яких містить шаблони триплетів. Планування запиту здійснюється окремо для кожного FBGP. Алгоритм пошуку релевантних джерел даних для запиту базується на простому порівнянні на відповідність шаблону триплета з доступними шаблонами триплетів в джерелах даних. Відповідність визначається рівністю предиката в шаблоні триплета предикату в джерелі даних, а також відповідністю обмежень на суб'єкт і об'єкт.

Нехай BGP становить множину шаблонів триплетів в FBGP. Результатом вибору джерела даних буде множина джерел даних D_j для кожного шаблону триплетів $t_j = (s_j, p_j, o_j) \in BGP$, де $D_j = \{d | (d, C) \in R \wedge$

$$\exists(p_j, r) \in C: r(s_j, o_j) = \text{true}\}$$

Побудова підзапитів. Результати вибору джерел використовуються для побудови підзапитів, на які кожне джерело може дати відповідь. Підзапити повинні складатися з одного FBGB на одне джерело даних. Подамо підзапит як триплет (T, C, d) , де T – множина шаблонів триплетів, C – множина обмежень значень, d – джерело даних, яке може відповісти на підзапит.

Оптимізація запитів. За результатами планування запиту план виконання запиту складається з декількох підзапитів. Задача оптимізатора запиту полягає в тому, щоб побудувати виконуваний і ефективний план виконання запиту, який враховує обмеження на шаблони доступу. Для побудови таких запитів використовується логічна і фізична оптимізація запиту.

Логічна оптимізація. Ця оптимізація запиту використовує еквівалентність виразів запиту для перетворення логічного плану запиту в еквівалентний план, який буде виконаний швидше або з меншими витратами. Рішення, запропоноване D2RQ, здійснює логічну оптимізацію двома шляхами. Спочатку, використовуються правила, описані у праці [61] для переписування оригінального запиту перед плануванням запиту таким чином, щоб об'єднати BGP там, де це можливо і де це можливо змінні замінюються константами із виразів з фільтрами.

Потім можливі обмеження значень переміщуються в підзапити для зменшення кількості проміжних результатів настільки рано, наскільки це можливо. Нехай $Q = (T, C, d)$ буде підзапитом і $FBGP = (T', C')$ – фільтрований базовий графовий шаблон. Обмеження значень C' може бути переміщене під запит, якщо всі змінні цього обмеження також використовуються в шаблонах триплета в підзапиті. Фільтри, які містять змінні із більш ніж одного підзапиту і які не можуть бути розділені з використанням обмеженої множини правил, повинні виконуватися локально в системі федеративних запитів.

Фізична оптимізація. Ця оптимізація запиту призначена для пошуку

«кращого» плану виконання запиту серед всіх доступних планів і використовує модель вартості для порівняння різних планів. Вартість запиту визначається на основі інформації щодо кількості доступних в кожному джерелі даних.

Виконання запиту. На цьому етапі можна широко застосовувати багатопоточність: скільки буде підзапитів до різних джерел даних, стільки буде виділено багатопоточних процесів. Розпаралелювання забезпечується технологіями Java Multithreading.

Об'єднання результатів запиту. Тут можна використовувати відомі результати. Дещо окремим питанням залишається реалізація *Nested Loop Join*.

Але навіть виконання цього етапу користувачем власними засобами буде виправданим, оскільки доцільність об'єднання результатів запиту не завжди легко визначити, оскільки дані дуже різноманітні.

3.2. Моделі міжагентної взаємодії

Розпочнемо розгляд з системного рівня подання мультиагентної системи та структурної організації її агентів. Агент представляється у вигляді набору елементів, що будують граф із можливими шляхами обробки вхідних заявок. Запропоновано механізм обробки заявок агентом, який надає можливість адаптації системи до класу задач, що розв'язуються, та динамічної зміни цілей агентів. У цьому розділі представлено модель організації МАС з адаптивними агентами, запропоновану автором в статті [9].

Прогрес у розвитку засобів обчислювальної техніки забезпечує можливість розв'язувати все більш широкий клас задач за допомогою програмних систем. Це обумовлює необхідність застосування нових підходів до їх проектування та розробки. Один із таких підходів оснований на застосуванні ідеї колективної діяльності сукупності агентів – мультиагентних систем (МАС) [9].

Одним із успішних застосунків МАС стало агентно-орієнтоване моделювання, яке надало можливості створення адекватних моделей у таких

сферах, як: біологія, соціологія, машинобудування та ін. Зокрема, представивши у вигляді агентів вірус імунodefіциту людини та лімфоцити, за допомогою МАС у роботі [111] моделювався перебіг захворювання ВІЛ-інфікованих з метою визначення причин значної диференціації проміжків часу між зараженням та проявом СНІД у різних хворих. У роботі [112] за допомогою МАС була змодельована система контролю за рухом повітряного транспорту. Варто відзначити, що в подібних системах поведінка кожного з агентів точно визначається відповідно до об'єктів, які за допомогою них моделюються.

Іншим напрямком розвитку МАС є вироблення концепцій агентно-орієнтованого проектування (АОП) програмних систем, яке може використовуватися у сферах електронної комерції, управління та моніторингу телекомунікаційних систем, оптимізації транспортних систем чи промислового виробництва, аналізу інформації та інформаційних середовищ (наприклад, Інтернет), автоматичного планування зустрічей, електронних розваг чи інтерактивних ігор [113].

Поклавши в основу розробки платформи підтримки міждисциплінарних досліджень мультиагентний підхід, перейдемо до розробки моделей для реалізації агентів, організації МАС з агентами, здатними адаптуватися до динамічної зміни та цілей агента.

Фундаментальними ознаками мультиагентної комп'ютерної системи приймаємо її здатність виконувати у певному середовищі автономні дії з метою досягнення поставлених під час проектування цілей [114] і включення набору автономних агентів, що можуть взаємодіяти між собою [115]. Такі системи використовуються для розв'язання задач, рішення яких є неможливим або занадто складним при використанні одного агента чи монолітної системи. Глобальна функція мультиагентної системи не задається чи проектується, вона виникає на основі взаємодії агентів між собою в наслідок синергетичного ефекту.

При розробленні мультиагентної системи використовуватимемо такі загальноновизнані їх властивості, як автономність, соціальність, реактивність та проактивність, оптимальність, адаптивність. Це обумовлює необхідність включення в агента засобів для оцінювання поведінки та продуктивності, планування і перепланування дій згідно визначених цілей [116].

Мультиагентний підхід забезпечує придатність для подальшого розширення функціональності шляхом впровадження агентів, що забезпечують певні перетворення даних, інтеграцію даних відповідно до їх семантично однорідних фрагментів і візуалізацію у вигляді карт, графіків і інших уявлень.

Для реалізації MAC була обрана середовище JADE (Java Agent DEvelopment Framework), яке відповідає специфікації FIPA (Foundation for Intelligent Physical Agents) [117], яка є стандартом IEEE Computer Society. Розроблена в дисертаційній роботі архітектура передбачає існування $n \geq 1$ агентів, які обслуговують джерело даних і мають унікальні для їх предметної області OWL онтології і файли XML, що визначають параметри підключення до джерела, правила проектування словника на джерело даних і параметри функціонування агента в інтегрованій системі.

Кожен такий агент періодично або опитує спеціалізованого агента – «дошку оголошень», або отримує повідомлення про наявність нових запитів користувача, виконує їх і формує результат. Користувач взаємодіє з системою через web-інтерфейс, за допомогою спеціалізованого агента, що організує зв'язок web-сервера з агентним середовищем. Агенти можуть звертатися до джерел як локально так і віддалено, якщо це дозволяє специфіка доступу до джерела. Агенти мають можливість мігрувати, щоб оптимізувати логічне виведення над словником даних при обробленні запиту користувача і розподілити навантаження. Пошук агентів всередині MAC і сервісів, які вони надають, здійснюється за спеціальним каталогом DF (Directory Facilitator), функціонування якого забезпечує Агентне середовище. Також в системі передбачено впровадження агентів, які забезпечують перетворення отриманих

результатів на запит користувача і агентів, що забезпечують перетворення запиту на підмножині природної мови в запит SPARQL (SPARQL Protocol and RDF Query Language).

На рис. 3.3 наведено приклад взаємодії агентів в системі, де відображені акти комунікації для виконання реєстрації агентів в DF, опитування джерел даних про їх належність до розділів Global Change Master Directory (GCMD) Science Keywords та внесення відповідних знань в OWL агента-класифікатора, відповіді на запит користувача.

В умовах підтримки за рахунок міжагентної взаємодії інтеграції застосунків і джерел даних засоби, що забезпечують визначення семантики даних, виходять на перший план. Для визначення семантики в роботі вибраний підхід, який активно розвивається і орієнтований на широку сферу застосування мов визначення семантики даних – мова WEB-онтологій – Web Ontology Language (OWL) [118]. Цей підхід запропонований і розвивається W3C.

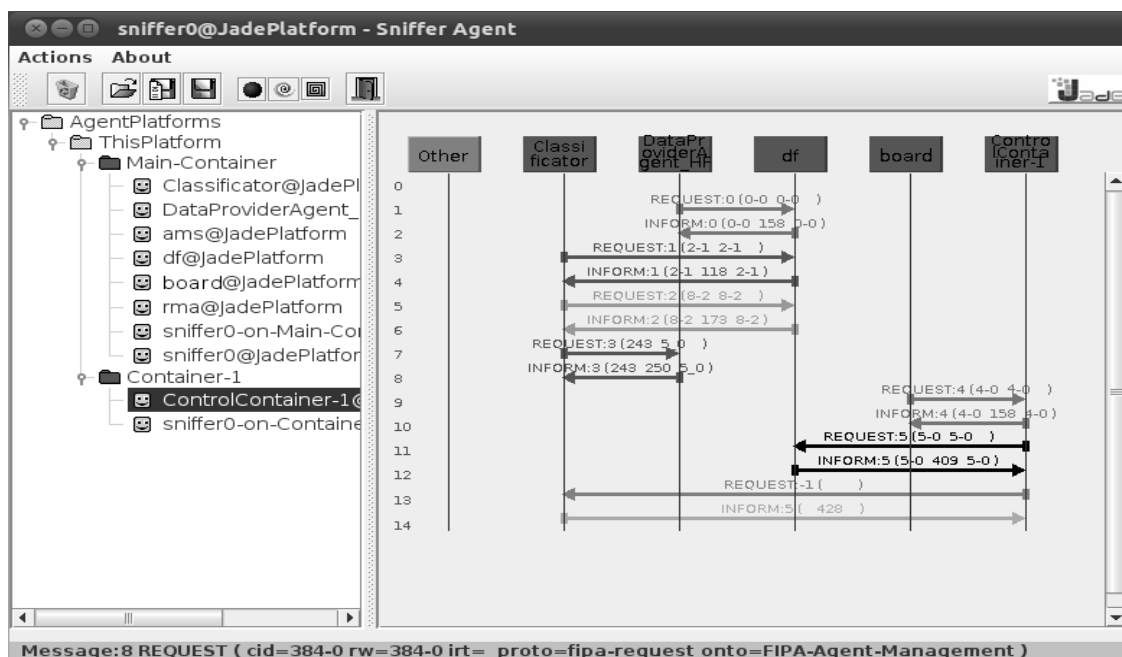


Рис.3.3. Приклад взаємодії агентів в системі

Парадигма Linked Data забезпечує інтероперабельність даних і можливість виділення семантично однорідної інформації. Вона передбачає повсюдні посилання – в спеціалізованих словниках на загальновизнані або

родинні ресурси за допомогою URI. Наприклад, масиви інформації з двох джерел, позначені в OWL, за допомогою властивості словника OWL `equivalentProperty` або іншого як <http://dbpedia.org/property/year> [119], незалежно від фізичного представлення («1999», «99»), дає можливість стверджувати її семантичну приналежність до року, а значить використовувати для їх об'єднання якийсь сервіс, орієнтований на перетворення і зіставлення дат.

Для зіставлення класів і властивостей, визначених у словнику OWL Tbox (terminological component) їх екземплярів в джерелі даних Abox (assertion component) використаний інструмент, що забезпечує проектування відповідних компонент – D2RQ [109]. Зв'язок з джерелом даних здійснюється за допомогою jdbc драйверів, які існують практично для будь-яких типів зберігання інформації – від БД до плоских файлів і html сторінок з таблицями.

Запити користувача до системи формуються у вигляді SPARQL [59] запитів, затвердженому W3C як стандарт запитів до RDF даних. Для зниження порогу входження користувачів системи, використовується парадигма фасетного пошуку [120]. При цьому фасети автоматично складаються на підставі словника, визначеного для кожного джерела даних.

Системний рівень організації МАС. Складовими частинами мультиагентної системи є множина агентів та їхнє середовище. Оскільки така система не має інших складових, необхідний зв'язок з іншими системами здійснюється через агентів. Кожен із агентів здатен аналізувати стан середовища, у якому знаходиться, та генерувати певний вплив на це середовище. Ці входні та вихідні дані агента A мають свої алфавіти (X та Y відповідно), а сам агент може розглядатися як перетворення над даними в цих алфавітах $A: X \rightarrow Y$.

Дані на вході агента можуть містити інформацію про входні аргументи його перетворення з алфавітом X_D та про директиви контролю цього перетворення з боку середовища з алфавітом X_C . Тобто, $X = X_D \times X_C$.

Насправді, вхідний алфавіт агента характеризує дані, у яких він може бути зацікавлений, вихідний – дані, які він може надати системі. Це означає, що факт появи даних на вході агента ніяким чином не гарантує появу даних на його виході. Однак, якщо агент має вихідний алфавіт, то вважається що в процесі роботи системи він обов'язково спродукує дані (варто відзначити, що тривалість роботи системи ніяк не обмежується), і в контексті всієї тривалості роботи системи агент здійснюватиме перетворення. Справедливе і протилежне твердження: якщо агент не надає відомостей про свій вихідний алфавіт, то він здатен лише отримувати вхідні дані. Для таких агентів вихідний алфавіт має вигляд $O = \{\emptyset\}$, тобто можна записати тільки пусті дані. Таким чином, можливі три типи агентів [9]:

- 1) агенти-перетворювачі (з повними вхідним та вихідним алфавітом),

$$A: X \rightarrow Y;$$

- 2) агенти-споживачі (з пустим алфавітом вихідних даних),

$$A: X \rightarrow O;$$

- 3) агенти-генератори (з пустим алфавітом вхідних аргументів),

$$A: O \times X_c \rightarrow Y.$$

Існування агентів останніх двох типів обґрунтовується необхідністю зв'язку з іншими зовнішніми системами. Далі розглядатимуться агенти першого типу.

Поставимо умову, що сенсори агентів отримують від середовища дані, представлені кортежем $R = \langle X', Y', x \rangle$, який несе інформацію про заявку на виконання перетворення $X' \rightarrow Y'$ над аргументами x . При цьому, оскільки агент є автономним, він здатен самостійно зробити запит типу R до середовища.

Щоб називатися адаптивною та автономною, система повинна володіти механізмом самостійної оцінки власного стану та винесення рішення про необхідність змін у роботі. Для здійснення такої оцінки у системі існує p параметрів роботи агента. У такому випадку стан агента може описуватися вектором $\bar{s} \in \mathbb{R}^p$, який містить значення кожного з параметрів.

Структурний рівень організації агента. Нехай агент A визначається набором елементів $a_i, i = \overline{1, n}$, кожен з яких може виконувати перетворення f_{a_i} над підмножиною вхідного алфавіту агента X^i . При цьому результатом такого перетворення є підмножина вихідного алфавіту агента Y^i . Тобто,

$$A = \{a_i | f_{a_i} : X^i \rightarrow Y^i\}, i = \overline{1, n},$$

$$X = \bigcup_i X^i, Y = \bigcup_i Y^i.$$

Семантика кожного такого перетворення визначається парою вхідного та вихідного алфавітів та символом перетворення (X^i, Y^i, f_{a_i}) . Дані про елементи, а саме про їхній вхідний та вихідний алфавіт, заносяться до внутрішнього реєстру агента.

Вважатимемо, що набір елементів в агенті може змінюватися завдяки реконфігурації. При цьому поява нового елемента може означати для агента або розширення його функціональності, або появу альтернативних засобів оброблення даних.

Знання про внутрішню структуру агента можуть бути представлені у вигляді орієнтованого графа G з множиною вершин V та множиною дуг E . Множина перетворень, які здійснюють елементи агента, є підмножиною прямого добутку його вхідного та вихідного алфавітів:

$$A \subseteq X \times Y = \bigcup_i X^i \times \bigcup_i Y^i.$$

Зазначена множина елементів агента складає множину вершин графа G , тобто $A=V$. Множина дуг є підмножиною усіх відношень на множині A та визначається як

$$E = \{(j, k) : Y^j \subseteq X^k\};$$

$$j = \overline{1, n}; k = \overline{1, n}.$$

При появі на вході агента заявки $R = \langle X', Y', x \rangle$, він може відреагувати на неї, виконавши відповідне перетворення. При цьому необхідною умовою оброблення є

$$\begin{cases} X' \subseteq X, \\ Y' \subseteq Y; \end{cases} \Rightarrow \exists \{j|j \in [1, n]\}, \{k|k \in [1, n]\}: \\ X' = \bigcup_{\{j\}} X^j \cap Y' = \bigcup_{\{k\}} Y^k.$$

Виконання даної умови означає, що існують такі елементи агента, вхідний та вихідний алфавіти яких відповідають заявці, що надійшла.

Достатньою умовою виконання заявки агентом є існування не пустої множини шляхів $\{P_{j,k}\}$ між відповідними вершинами j та k :

$$P_{j,k} = \{p_{j,k}\} = \langle a_j, \dots, a_k \rangle.$$

При цьому вважатимемо, що передача даних по одній із дуг графа G пов'язана зі зміною стану агента (Δs). Тобто існує множина функцій $\{\varphi: E \rightarrow R^p\}$, які обчислюють вартість кожного з переходів у графі для агента.

Таким чином, структуру S агента можна описати через множину його елементів-перетворень A , вектор стану \bar{s} та множину функцій, які визначають вартість кожного з переходів по графу G :

$$S = \langle A, \bar{s}, \{\varphi\} \rangle.$$

Граф G може мати декілька маршрутів між будь-якими парами вершин j та k , що дає можливість зміни шляху оброблення заявки в залежності від вхідних даних x , поточного набору елементів $A = \{a_i\}$, та значень параметрів стану агента \bar{s} . При цьому потрібно розв'язувати задачу оптимізації.

Процес обробки заявки в агенті. Із появою заявки R на вході агента у випадку, коли вона може бути оброблена, в агенті починається просування пакетів з даними. Даний пакет C у момент часу τ є четвіркою, яка складається з унікального ідентифікатору поточного запиту $u \in N$, алфавіту даних, які обробляються M , поточних даних d для обробки елементом агента та шляху P , який даний пакет пройшов в агенті (упорядкована підмножина елементів останнього):

$$C^\tau = \langle u, M^\tau, d^\tau, P^\tau \rangle, P^\tau \subseteq A.$$

При $\tau = 0$ маємо $M^\tau = X', d^\tau = x, P^\tau = \{\}$.

Після оброблення пакета елементом a_t ($t \in [1, n]$) агента, частини M та d змінюються відповідно до вихідного алфавіту поточного елемента та результатів здійсненого перетворення:

$$\begin{cases} M^{\tau+1} = Y^t, \\ d^{\tau+1} = f_{a_t}(d^{\tau}), \\ P^{\tau+1} = P^{\tau} \cup a_t. \end{cases}$$

Далі відбувається передача пакета $S^{\tau+1}$ наступним елементам. На агента покладається задача визначення, кому з елементів необхідно передати заданий пакет. Останній може бути доставлений лише тому елементу, який має відповідний вхідний алфавіт. Отже, при прийнятті рішення про пункт призначення пакета агент формує вибірку елементів із множини доступних (тих, які можуть обробити поточні дані), керуючись результатами аналізу поточного стану системи. Така вибірка розміром з m елементів може бути описана наступним чином:

$$\{a_i | M^{\tau} \subseteq X^i\} \subseteq A, i = \overline{1, m}.$$

Слід вказати, що ситуація з декількома маршрутами можлива, коли агент має в своєму складі елементи, чиї вхідні алфавіти перетинаються:

$$m > 1 \Rightarrow \exists j, k \in [1, n]: X^j \cap X^k \neq \{ \}.$$

Зміни у роботі системи можуть бути здійснені шляхом модифікації рішень стосовно доставки пакета через граф G агента.

Агент повинен вибрати маршрут з найменшою вартістю для просування пакету. Тоді, коли такий маршрут невідомий, виконується широкомовна передача пакета елементам, які можуть обробити дані, що передаються всередині. Із усіх можливих маршрутів, утворених таким чином, вибирають ті, які дають мінімальний час проходження пакета. Такі маршрути запам'ятовуються та використовуються для подальших обробок відповідних заявок.

Виходячи зі свого поточного стану та набору функцій φ , агент може прийняти рішення про «небажаність» проходження пакету по певній дузі

графа G або маршруту. У такому випадку агент може уповільнювати проходження даних по вибраній дузі і корегувати таким чином вибір маршруту.

Отже, агент отримує можливість зміни способів оброблення вхідних заявок у залежності від власних цілей, які визначаються функцією φ вартості. Така реконфігурація маршруту здійснюється у випадку появи в структурі агента елемента, який може входити до альтернативного маршруту, або зміни вартості маршрутів обробки заявки.

3.3. Системне узгодження даних різної природи

Використовувані в міждисциплінарних дослідженнях дані мають різну природу, яка визначається їх об'єктивним змістом, цільовим призначенням і способом отримання [121, 122]. В цьому випадку побудова агрегованих міждисциплінарних моделей [8, 13, 123, 124] вимагає приведення цих даних до єдиної семантики, єдиного діапазону значень і єдиних одиниць виміру за умови мінімізації інформаційних втрат такого узгодження.

Узгодження даних породжує необхідність вирішення таких завдань:

- 1) оцінка інформаційних втрат і зростання невизначеності моделі, що виникають в результаті узгодження даних.
- 2) розроблення методології кількісної оцінки узгодженості даних різної природи,
- 3) розроблення алгоритмів і методів такого узгодження.

У загальному випадку, дослідження деякого явища пов'язано з аналізом і обробленням відомостей про нього. Ці відомості є кількісні і (або) якісні оцінки властивостей деякої сукупності об'єктів $O = \{o_i\}, i = \overline{1, n}$, де o_i – значення ідентифікаторів номінальної шкали об'єктів представленої сукупності. Можна говорити, що такі дані є результатом відображень вигляду:

$$O \xrightarrow{I_j} X^j, j = \overline{1, m} \quad (3.1)$$

де $I_j, j = \overline{1, m}$ – відображення, визначені на множині об'єктів ($D(I_j) = O$). Їх області значень – це області визначення показників X^j , тобто $E(I_j) = D(X^j)$.

Семантика відображень $I_j, j = \overline{1, m}$ і ототожнюваних з ними показників у (3.1) формулюється, виходячи з цілей дослідження, і визначає об'єктивний зміст (яка властивість оцінюється), цільове призначення (для чого призначені ці оцінки) і спосіб отримання (дані можуть бути результатом вимірювань, моделювання, експертного оцінювання).

Інформаційні втрати при перетворенні вимірювальних шкал. Можливості спільного використання даних різної природи залежать від типів шкал, в яких вони виміряні [125]. Мета вимірювальних експериментів полягає у визначенні стану емпіричної системи $E = (O, R)$, де $R = \{r_k\}, k = \overline{1, m}$ – множина відношень на множині O , за допомогою поставленої їй у відповідність вимірювальної системи $M = (O', R')$, де $O' = \{o'_k\}, k = \overline{1, n'}$ – множина символів, $R' = \{r'_l\}, l = \overline{1, m'}$ – множина допустимих відношень на них. Відповідність між емпіричною і вимірювальною системами задається сюр'єкцією $g: E \rightarrow M$.

Шкала вимірювань являє собою трійку $S = (E, M, g)$, яка повністю визначає процес вимірювання [126]. Оцінка інформаційних втрат при вимірах у цій шкалі пов'язана з оцінкою невизначеності зворотного відображення $g^{-1}: M \rightarrow E$. Інформація, отримана в результаті безпосереднього з'ясування стану емпіричної системи, визначається як $I_E = H(E)$, де $H(E)$ – її власна ентропія [127]. Але, як правило, стан емпіричної системи можна визначити лише опосередковано з допомогою шкали $S = (E, M, g)$. В цьому випадку емпірична і вимірювальна системи розглядаються як залежні, і має місце співвідношення:

$$I_{M \rightarrow E} = H(E) - H(E | M) \quad (3.2)$$

де $H(E|M)$ – умовна ентропія, що характеризує невизначеність стану емпіричної системи у випадку, а стан вимірювальної системи визначений повно.

Таким чином, інформаційні втрати можна оцінити за допомогою величини.

$$\Delta I(M, E) = I_E - I_{M \rightarrow E} = H(E|I). \quad (3.3)$$

Повна інформація систем E і M , визначена за виразом (3.2), є симетричною:

$$H(E) - H(E|M) = I_{M \rightarrow E} = I_{E \rightarrow M} = H(M) - H(M|E), \quad (3.4)$$

де $H(M|E)$ – умовна ентропія вимірювальної системи.

У разі відсутності побічного впливу на вимірювальну систему її вважають підпорядкованою емпіричній системі ($H(M|E)=0$), а вираз (3.3) записують як:

$$\Delta I(M, E) = H(E) - H(M). \quad (3.5)$$

Процес перетворення даних із шкали S_1 в шкалу S_2 розглядаємо як процес вимірювання, в якому S_1 відіграє роль емпіричної системи, а S_1 – вимірювальної, тобто можна визначити нову шкалу $T_{S_1 \rightarrow S_2} = (S_1, S_2, \varphi)$, де $\varphi: S_1 \rightarrow S_2$.

Шкали S_1 і S_2 будемо вважати еквівалентними, якщо для $\exists \varphi \in \Phi: S_1 \rightarrow S_2$ має місце $\Delta I(S_1, S_2) = 0$. Зрозуміло, що відносно класів $\Phi = \{\varphi\}$ множина шкал $T_{S_1 \rightarrow S_2} = (S_1, S_2, \varphi)$ поділяється на класи еквівалентності (типи) S_Φ і перетворення даних в межах шкал одного типу не призводить до інформаційних втрат. Якщо визначити $O = X^i \subseteq R, O' = X^j \subseteq R$, то прикладами таких типів шкал будуть [126]:

- кількісні шкали, для яких

$$\Phi = \{\varphi: x_i = ax_j + b\}, \quad S_i, S_j \in S_\Phi, \quad x_i \in X^i, x_j \in X^j,$$

a, b – параметри масштабу і зміщення;

- порядкові шкали, для яких

$$\Phi = \{ \varphi : \forall x_i \leq x_j \Rightarrow \varphi(x_i) \leq \varphi(x_j) \}, \quad S_i, S_j \in S_{\Phi}, \quad x_i, x_j \in X^i, \varphi(x_i), \varphi(x_j) \in X^j;$$

- номінальні шкали, для яких

$$\Phi = \{ \varphi : \forall x_i \neq x_j \Rightarrow \varphi(x_i) \neq \varphi(x_j) \}, \quad S_i, S_j \in S_{\Phi}, \quad x_i, x_j \in X_i, \varphi(x_i), \varphi(x_j) \in X_j.$$

Ці типи шкал є найбільш поширеними і дозволяють давати як кількісну, так і якісну оцінку властивостей досліджуваних об'єктів. Таким чином, рішення задачі узгодження даних зводиться до побудови процедури перетворення $\varphi : S_1 \rightarrow S_2$. При цьому шкали S_1 і S_2 можуть належати до одного або різних типів. Якщо шкали S_1 і S_2 різних типів і $\Delta I(S_1, S_2) = H(S_1) - H(S_2) > 0$, то шкала S_1 «сильніша», ніж шкала S_2 , і при переході $S_1 \rightarrow S_2$ виникають інформаційні втрати. При зворотному ж переході, коли $\Delta I(S_2, S_1) < 0$, в модель вноситься невизначеність оцінювана числом $-\Delta I(S_2, S_1)$.

Серед відомих шкал найбільш «слабкими» є номінальні шкали, найбільш «сильними» – кількісні шкали, порядкові шкали займають проміжне положення [128].

Вираз для кількісного оцінювання інформаційних втрат, що виникають при переході від кількісної шкали до порядкової або номінальної, формулюється в термінах кількості відліків n вихідної (кількісної) шкали і кількості відліків m цільової (порядкової або номінальної) шкали. Цей вираз для порядкової шкали визначається наступним чином:

$$\Delta I(n, m) = \sum_{l=1}^m \frac{R(n, l) C_m^l \log(C_{\max(n, l)}^{\min(l, n)})}{m^m}, \quad (3.6)$$

а для номінальної шкали, набуває вигляду:

$$\Delta I(n, m) = \sum_{l=1}^m \frac{S(n, l) A_m^l \log(A_{\max(n, l)}^{\min(l, n)})}{m^m}. \quad (3.7)$$

У виразах (3.6) і (3.7) C_m^l , $C_{max(n,l)}^{min(l,n)}$ означають число сполучень, A_m^l і

$A_{max(n,l)}^{min(l,n)}$ – число розташувань, $R(n,l) = \sum_{j=0}^l (-1)^j C_l^j (l-j)^n$ – число Моргана, а

$S(n,l) = \frac{R(n,l)}{l!}$ – число Стірлінга другого роду.

Таким чином, при перетворенні даних з метою їх узгодження слід враховувати можливі інформаційні втрати або зростання невизначеності моделі, які виникають при переході від одних типів шкал до інших.

Невизначеність моделі, що виникає при переході від «слабкіших» шкал до «сильнішим», може бути частково зменшена за рахунок використання додаткової інформації про досліджувані об'єкти. Наприклад, якщо при кластеризації об'єктів за будь-яким показником X^j зберегти інформацію про середні значення для кожного кластера C^j , то використання цієї додаткової інформації при зворотному переході від номінальних значень кластерів C^j до кількісних оцінок X^j дозволяє знизити невизначеність моделі.

Корисність використання додаткової інформації при перетворенні даних можна оцінити за допомогою формули:

$$U(I, S_1, S_2) = \Delta I(S_1, S_2) - \Delta I((S_1, S_2) | I), \quad (3.8)$$

де $U(I, S_1, S_2)$ – корисність додаткової інформації I при переході від шкали S_1 до шкали S_2 , $\Delta I(S_1, S_2)$ і $\Delta I((S_1, S_2) | I)$ – втрати інформації при перетворенні даних відповідно без використання і з використанням додаткової інформації I .

На рис. 3.4. представлена експериментально отримана залежність між довжиною протоколу вимірювань k (кількість об'єктів, що підлягають аналізу), і нормованими значеннями $\Delta I(S_1, S_2)$, $\Delta I((S_1, S_2) | I)$ і $U(I, S_1, S_2)$ для наведеного вище прикладу.

Як видно з рис. 3.4, саме для критичних довжин протоколів, коли інформаційні втрати досягають максимуму, корисність додаткової інформації

виявляється найбільшою і складає в відносних одиницях близько 60% від максимального значення інформаційних втрат

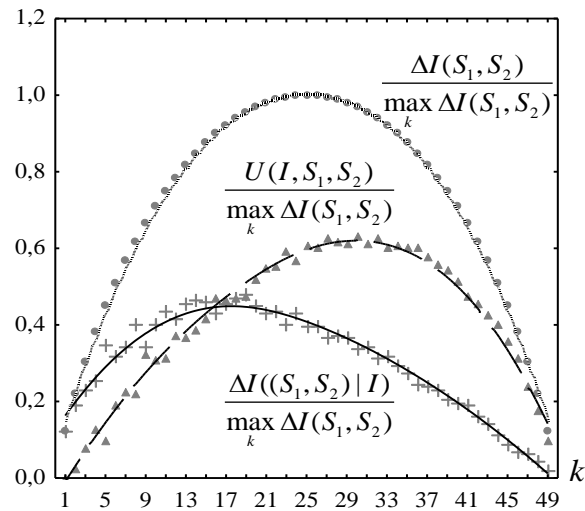


Рис. 3.4. Залежність інформаційних втрат від довжини протоколу вимірювань для переходу від номінальної шкали до порядкової шкали.

Побудова метрик для оцінки узгодженості даних. Усі наявні в розпорядженні дослідника дані вигляду (4.1) можуть бути представлені у вигляді таблиці «об'єкт-властивість» [121]:

$$X = (x_{i,j})_{i=1,j=1}^{n,m}, \quad (3.9)$$

в якій рядок X_i відповідає набору значень, що характеризує властивості об'єкта o_i , а стовпець X^j задає значення j -ого показника для всієї вибірки об'єктів.

Визначені в термінах даних різні погляди на одне і те ж явище, такі як «вибірки об'єктів», «часові ряди», «проекції властивостей» і т.п., відповідають різним підмножинам $V \subseteq X$, які можуть бути задані як композиції на множині елементарних вибірок $V_i \equiv X_i$ – горизонтальних перетинів таблиці (3.9), і елементарних проекцій $V^j \equiv X^j$ – вертикальних перетинів тієї ж таблиці. Якщо позначити як \tilde{V} довільну вибірку або проекцію, а як \tilde{X} – елементарну вибірку або проекцію, то отримуємо поділ \tilde{V} по множині \tilde{X} :

$$\tilde{V} \subseteq \bigcup_{i=1}^{|\tilde{X}|} \tilde{X}_i \quad 3.10$$

Підхід до побудови метричних просторів для кількісної оцінки узгодженості даних, докладно викладений в роботі [129], ґрунтується на використанні поділу (3.10).

У загальному випадку на множині $\tilde{X} = \{\tilde{X}_i\}$ можна задати сигма-алгебру \mathbb{S} [130] і визначити метричний простір $(\tilde{X}, \mathbb{S}, \mu)$, ізоморфний ймовірнісному простору. Тобто незалежно від типів шкал представлення даних відбувається відображення їх значень в номінальні події, визначені на множині \tilde{X} .

Якщо дані представлені в кількісних шкалах, $\tilde{X}_i \in \tilde{X}$ можна розглядати як елементи лінійного простору $E(X^1) \times \dots \times E(X^m)$ або використовувати інші моделі, наприклад кватерніони [131], на яких введена метрика.

У тому випадку, коли вдається визначити простір з нормою:

$$\|\tilde{X}_i\| = \left(\sum_{j=1}^{|\tilde{X}|} w_j (x_{i,j})^p \right)^{\frac{1}{p}}, \quad 3.11$$

де w_j – вагові коефіцієнти об'єктів або показників, з'являється можливість інтегральної оцінки поглядів за допомогою норми,

$$\|\tilde{V}_i\| = \left(\sum_{\tilde{X}_k \in \tilde{V}_i} \|\tilde{X}_k\|^p \right)^{\frac{1}{p}}$$

а також близькості таких поглядів:

$$\|\tilde{V}_i\| - \|\tilde{V}_j\| \leq d(\tilde{V}_i, \tilde{V}_j) = \|\tilde{V}_i - \tilde{V}_j\| \leq \|\tilde{V}_i\| + \|\tilde{V}_j\|. \quad (3.12)$$

При $p=2$, маємо евклідовий простір [132], що дає можливість визначити в ньому поняття ортогональності (незалежності) поглядів. Слід зауважити, що визначення відстані в співвідношенні (3.12) може варіюватися в залежності від цілей дослідження. Наприклад, замість $\|\tilde{V}_i - \tilde{V}_j\|$ часто використовують $\|\bar{V}_{i,j}\|$, де $\bar{V}_{i,j}$ – деяке усереднення (середнє або медіана)

[126], знайдене як рішення задачі $\operatorname{argmin}_{V \in \mathbb{S}} (d(\bar{V}_{i,j}, \tilde{V}_i) + d(\bar{V}_{i,j}, \tilde{V}_j))$. Можуть бути використані й інші норми, що найбільш точно відображають зміст «близкості» поглядів.

Так для оцінки «близкості» статистичних розподілів запропонована міра $L(P, Q) = 2H(\frac{P \oplus Q}{2}) - H(P) - H(Q)$ [129], де P і Q – оцінювані розподіли, $\frac{P \oplus Q}{2} = \left\langle \frac{p_1 + q_1}{2}, \frac{p_2 + q_2}{2}, \dots, \frac{p_n + q_n}{2} \right\rangle$ – усереднений розподіл, $H(\frac{P \oplus Q}{2})$, $H(P)$ і $H(Q)$ – ентропія Шеннона [133].

На основі цієї міри можна розробити цілий ряд оцінок, корисних для вирішення практичних задач. Так, в роботі [129] запропонований підхід, який полягає у визначенні найкращого і найгіршого з точки зору вирішення конкретної прикладної задачі розподілу. Тоді для оцінки можна використовувати вираз

$$S(P) = \frac{L(P, P_{\inf}) - L(P_{\inf}, P_{\inf})}{L(P_{\sup}, P_{\inf}) - L(P_{\inf}, P_{\inf})}. \quad (3.13)$$

Враховуючи, що $L(P_{\inf}, P_{\inf}) = 0$, вираз (3.13) можна записати у вигляді $S(P) = \frac{L(P, P_{\inf})}{L(P_{\sup}, P_{\inf})}$. Очевидно, що $S(P) \in \{0, 1\}$. Отримане значення можна прийняти за кількісну оцінку розподілу, виражену у відносній шкалі. Маючи множину розподілів $\pi = \{P_j, j = \overline{1, m}\}$ і визначивши граничні розподіли P_{\sup} і P_{\inf} , можемо визначений нестрогий порядок $\alpha \subseteq \pi \times \pi: \langle p_i, p_j \rangle \in \alpha \Leftrightarrow S(p_i) \geq S(p_j)$ і визначити бажаний критерій через вирішення оптимізаційних задач вигляду $\max_{P \in \pi} (S(P)), \arg \max_{P \in \pi} (S(P))$.

Методи перетворення даних. При розробленні і використанні методів перетворення даних необхідно враховувати обмеження, що накладаються типами шкал, в яких представлені дані. Так перетворення даних в межах одного типу шкал повинні забезпечувати відсутність інформаційних втрат. У

тому випадку, коли типи вихідної і цільової шкали не збігаються, необхідно мінімізувати такі втрати.

Якщо дані представлені в кількісних шкалах ($X^j \in R, j = \overline{1, m}$, де R – поле дійсних чисел [130]) і є необхідність приведення даних до єдиної семантики і діапазону значень, перетворення даних (значень показників) $C_j: X^j \xrightarrow{C_j} R$ має відбуватися без інформаційних втрат, що накладає умова монотонності $x_{k,j} \prec x_{l,j} \Leftrightarrow C_j(x_{k,j}) \leq C_j(x_{l,j}); k, l \in [1, n], k \neq l$. Цим вимогам відповідають лінійні і нелінійні нормування – монотонні параметричні перетворення.

Лінійне нормування визначається відповідно до виразу:

$$C_{norm}(x_{i,j}) = \frac{x_{i,j} - a}{b}, \quad (3.14)$$

де $x_{i,j}$ – значення з таблиці (4.9), a – параметр, що задає зміщення; b – параметр, що визначає масштаб нормування.

Прикладом нелінійного нормування може служити перетворення:

$$C_{norm}(x_{i,j}) = \left(1 - e^{-\frac{a - x_{i,j}}{b}} \right)^{-1}, \quad (3.15)$$

що задає логістичну криву [134], в якій параметри a і b мають таке ж значення, що і в (3.14). Параметри перетворень (3.14) і (3.15) вибирають таким чином, щоб привести вихідні дані до заданого діапазону значень (найчастіше до інтервалу $[0, 1]$). Множина параметрів таких нормувань повинні мати ті ж одиниці вимірювання, що і вихідні показники. У цьому випадку нормовані значення показників будуть безрозмірними, і ці значення слід трактувати як положення аналізованого об'єкта щодо деякої множини еталонів. Спосіб вибору таких еталонів визначатиме остаточно семантику нормованих значень. Часто для обчислення значень параметрів і використовують вирази:

$$a = \min_{i=1,n}(x_{i,j}), b = \max_{i=1,n}(x_{i,j}) - \min_{i=1,n}(x_{i,j}); \quad (3.16)$$

або

$$a = \overline{X^j} = \frac{1}{n} \sum_{i=1}^n x_{i,j}, b = \sigma(X^j) = \sqrt{\frac{\sum_{i=1}^n (x_{i,j} - \overline{X^j})^2}{n}}, \quad (3.17)$$

де $\overline{X^j}$ – середнє значення, а $\sigma(X^j)$ – стандартне відхилення показника X^j , що визначається на вибірці об'єктів O . В цьому випадку нормування проводиться щодо середнього по вибірці O .

Якщо в якості еталонів використовувати деякі об'єкти, які характеризуються граничними станами, то відображення (3.14) і (3.15) можуть інтерпретуватися як функції належності для лінгвістичних змінних [129]. Якщо показники задані в номінальних шкалах, спектр еквівалентних перетворень даних обмежується бієкціями. На рис. 3.5. наведена математична схема бієктивного узгодження даних, представлених в номінальних шкалах і в рамках якої завдання узгодження даних розглядається як задача оптимізації. За критерій оптимізації у цій схемі прийнята імовірнісна міра – альфа

$$\text{Кліппендорфа [135]: } \alpha_k = \frac{(n-1) \sum_{i=1}^n o_{i,i} - \sum_{i=1}^n s_i(s_i-1)}{n(n-1) - \sum_{i=1}^n s_i(s_i-1)}.$$

Значення α_k в діапазоні 0,75-1,0 відповідають високому ступеню, 0,5-0,75 – середньому ступеню, а значення менше 0,5 – низькому ступеню узгодженості даних, представлених в номінальних шкалах.

У праці [12] запропонований метод перетворення в порядкову шкалу даних, представлених в кількісних шкалах, за умови мінімізації інформаційних втрат, що виникають при цьому. З цією метою для показника X , вираженого в кількісній шкалі, необхідно побудувати варіаційний ряд $\tilde{X} = \langle \tilde{x} \rangle : \forall \tilde{x} \in \tilde{X}, \tilde{x} \in X, \forall \tilde{x}_i, \tilde{x}_j \in \tilde{X}, i < j \quad \tilde{x}_i \leq \tilde{x}_j$ і використовувати кусочно-лінійну апроксимацію його кумуляти, яка визначається наступним чином:

$C_k = \langle c_i \rangle : c_i = \sum_{l=1}^i \tilde{x}_l, i = \overline{1, n}$, де c_i – i -те значення кумуляти, $\tilde{x}_l \in \tilde{X}$ – l -тий член

варіаційного ряду, n – довжина цього ряду.

Визначення значень показника X в порядковій шкалі пов'язане з побудовою такого розбиття варіаційного ряду \tilde{X} на сегменти:

$$\pi(\tilde{X}) = \langle \langle \tilde{x}_1, \tilde{x}_2 \rangle, \langle \tilde{x}_2, \tilde{x}_3 \rangle, \dots, \langle \tilde{x}_{r-1}, \tilde{x}_r \rangle \rangle, \tilde{x}_l \in \tilde{X}, l = \overline{1, r}, \forall l_1, l_2 = \overline{1, r}, l_1 < l_2 : \tilde{x}_{l_1} \leq \tilde{x}_{l_2}$$

для якого виконується умова $\sum_{\substack{\tilde{x} \in \langle \tilde{x}_i, \tilde{x}_{i+1} \rangle \\ i = \overline{1, r-1}}} (\tilde{x} - M(\langle \tilde{x}_i, \tilde{x}_{i+1} \rangle))^2 \rightarrow \min$, де

$M(\langle \tilde{x}_i, \tilde{x}_{i+1} \rangle)$ – математичне очікування значень показника в сегменті варіаційного ряду. Таким чином, кожному сегменту $\langle \tilde{x}_i, \tilde{x}_{i+1} \rangle \in \pi(\tilde{X})$ ставиться у відповідність величина $M(\langle \tilde{x}_i, \tilde{x}_{i+1} \rangle)$, а кожній з них, у свою чергу, – значення ординальної шкали $X \rightarrow \tilde{X} \rightarrow \pi(\tilde{X}) \rightarrow \langle M(\langle \tilde{x}_i, \tilde{x}_{i+1} \rangle) \rangle, i = \overline{1, r-1} \rightarrow \overline{1, r-2}$.

Якщо ряд $\langle M(\langle \tilde{x}_i, \tilde{x}_{i+1} \rangle) \rangle$ використовується для зворотного перетворення, то невизначеність моделі буде описуватися залежністю, зображеною на рис. 3.5.

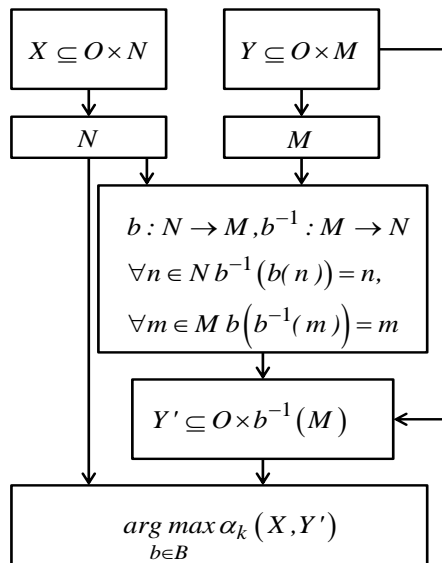


Рис. 3.5. Бієктивне узгодження даних, представлених в номінальних шкалах.

Висновки за розділом

1. Розроблено комплекс моделей і методів інтеграції джерел даних в рамках платформи підтримки міждисциплінарних досліджень у Світовій системі даних.
2. Розроблено моделі і методи федерування запиту на семантизовані дані, планування виконання запитів, оптимізації планів.
3. Для реалізації розробленої у розділі 2 архітектури МАС на системному рівні подання МАС виділені три типи агентів, запропоновані їхня структурна організація та процес оброблення заявок в них, які створюють можливість динамічної структурної реорганізації агентів з метою оптимізації процесу роботи МАС.
4. Оскільки побудова агрегованих міждисциплінарних моделей вимагає приведення цих даних до єдиної семантики, єдиного діапазону значень і єдиних одиниць виміру за умови мінімізації інформаційних втрат такого узгодження, у роботі запропоновані моделі системного узгодження даних різної природи, яке визначається їх об'єктивним змістом, цільовим призначенням і способом отримання.

РОЗДІЛ 4.

РОЗРОБЛЕННЯ ПЛАТФОРМИ ПІДТРИМКИ МІЖДИСЦИПЛІНАРНИХ ДОСЛІДЖЕНЬ І ЇЇ ЗАСТОСУВАННЯ В СВІТОВІЙ СИСТЕМІ ДАНИХ

У розділі описується створена платформа підтримки міждисциплінарних досліджень в Світовій системі даних, розглядаються питання її реалізації і особливості використання на прикладі побудови стратегій та сценаріїв сталого розвитку України. Розроблена платформа розглядається через призму Світової системи даних Міжнародної наукової ради. СЦД-Україна в рамках Світової системи даних щорічно виконує моделювання і аналіз сталого розвитку в глобальному та регіональному контекстах.

4.1. Архітектура платформи і принципи її функціонування

Розроблена платформа є розподіленою інформаційною системою з сервіс-орієнтованою архітектурою, призначеною для вирішення задач, що виникають в процесі реалізації безперервного життєвого циклу даних, які використовуються для наукових та прикладних досліджень. Платформа містить ресурси, що забезпечують моніторинг стану індикаторів, має розвинуті механізми аналізу даних та візуалізації. Набір засобів аналітичної обробки даних створено на основі сучасних інформаційних технологій, що забезпечує їх високу надійність та зручність у використанні. Кожен інструмент створено у вигляді окремого сервісу, що відповідає за роботу того чи іншого алгоритму обробки даних.

Призначення платформи в цілому визначає функціональність її підсистем та архітектурних блоків. Сервіс-орієнтована архітектура платформи підтримки міждисциплінарних досліджень зображена на рисунку 4.1. З огляду на призначення платформи, умови її функціонування в СЦД-Україна, перспективи використання в інших центрах ССД, до неї включено такі основні

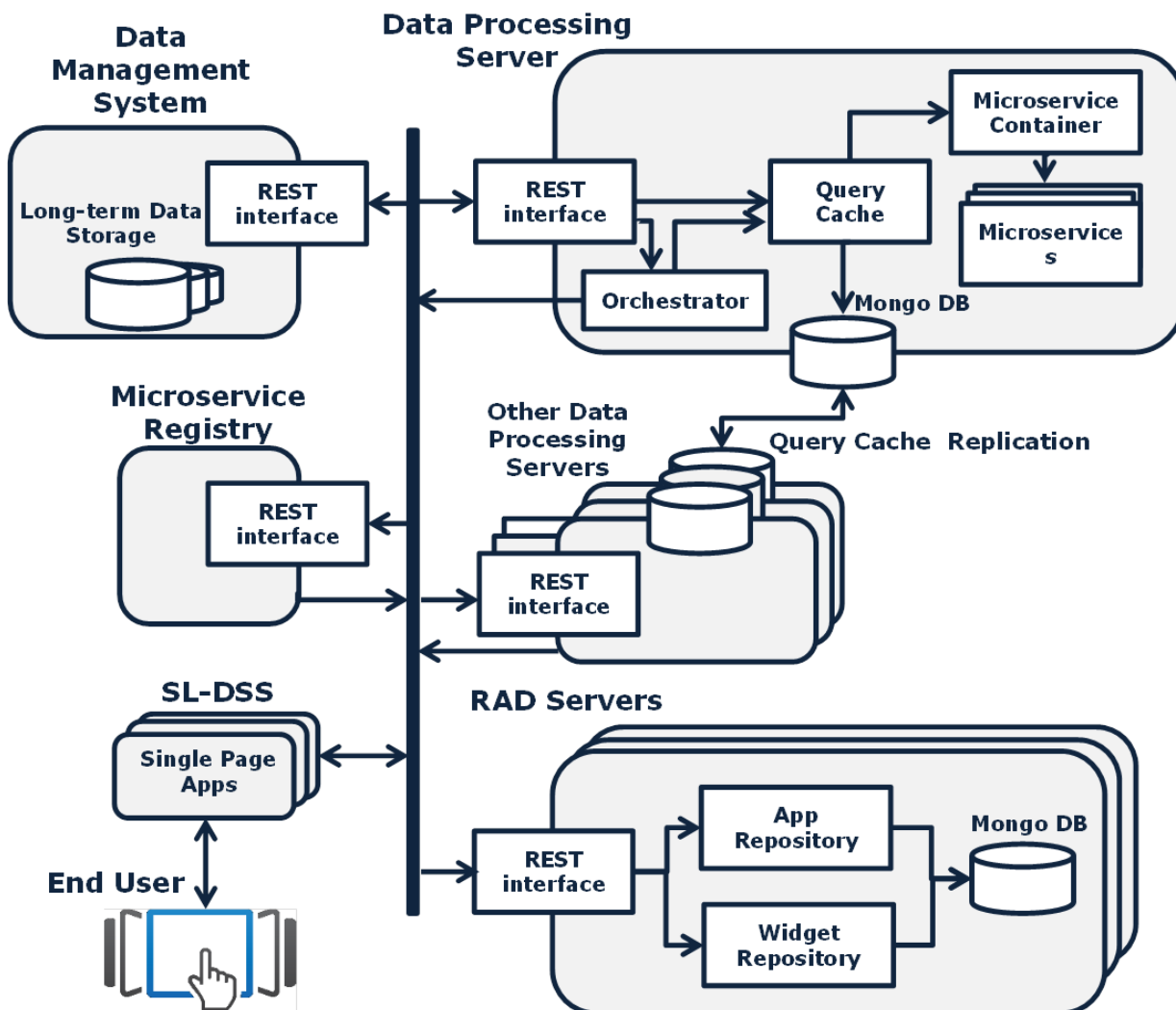


Рис. 4.1. Сервіс-орієнтована архітектура платформи підтримки міждисциплінарних досліджень

архітектурні компоненти:

- 1) підсистема довготривалого зберігання і управління даними (Data Management System), яка має функціональність, необхідну для вирішення задач управління даними;
- 2) підсистема опрацювання даних (Data Processing Server), яка реалізує алгоритми попередньої обробки, верифікації та аналітичного опрацювання даних;
- 3) підсистема швидкого розроблення і підтримки прикладних застосунків, що дозволяє вирішувати задачі розробки прикладних застосунків та управління ними;

4) загальний реєстр ресурсів, який дозволяє управляти конфігурацією розподіленої системи та здійснювати моніторинг її стану.

Як видно з рис. 4.1, функціонування платформи основане на взаємодії розподілених ресурсів декількох типів. Технічна реалізація платформи передбачає багатосерверну взаємодію (щонайменше $n100$ серверів оброблення даних і $n1000$ серверів швидкого розроблення застосунків), різноманіття систем збереження даних і засобів кінцевих користувачів.

Розглянемо призначення і особливості реалізації у складі платформи її основних компонентів.

4.1.1. Підсистема довготривалого зберігання і управління даними

Ця підсистема забезпечує базову функціональність ETL: публікацію і пошук наборів даних, їх зберігання та управління довготерміновими даними тощо. Пілотна версія була впроваджена в СЦД-Україна в декількох проектах.

Розглянемо *загальні принципи організації даних*. Дані, які зберігаються та оброблюються в підсистемах довготривалого зберігання і управління даними, подаються у вигляді наборів (datasets), що описують багатовимірні куби та супроводжуються метаданими, необхідними для визначення семантики цих даних.

Кінцевий користувач має змогу описувати дані у вигляді файлу в форматі .xlsx, який складається з аркушів, призначених для опису метаданих набору в цілому, власне даних, розширення метаданих (тлумачного словника) та засобів інтернаціоналізації.

Кожен з наборів ідентифікується за допомогою універсального унікального ідентифікатора (UUID), що забезпечує можливість міграції наборів даних між окремими вузлами підсистеми довготривалого збереження. Підсистема надає можливість контролю версій даних завдяки збереженню управління ланцюжками версій (commits), які супроводжуються коментарями адміністратора даних або генерованих засобами автоматичного оновлення даних повідомленнями.

Метадані, що супроводжують набір, подані деревоподібною ієрархічною структурою, яку можна описати за допомогою ланцюгів шляхів у формі <root>.<childNodes>....<valueNode>. При цьому для позначення піддерева використовують символ «*», а для позначення списку значень – «[]». В таблиці 4.1 зведено призначення окремих підструктур метаданих.

Таблиця 4.1.

Призначення підструктур метаданих

Підструктура метаданих	Призначення	Примітка
dataset.*	Опис набору даних	
dataset.id	Універсальний унікальний ідентифікатор набору	Попередньо генерується підсистемою довготривалого зберігання
dataset.status	Стан набору	опублікований, приватний
dataset.locale[]	Наявні локалізації	
dataset.label	Назва набору	
dataset.note	Опис набору	
dataset.source	Джерело даних	
dataset.topics[]	Ключові слова	Перелічуються через кому, можуть бути шляхами в ієрархії
dataset.commit.*	Опис версії набору	
dataset.commit.id	Внутрішній ідентифікатор версії	Генерується автоматично
dataset.commit.author	Автор версії	Генерується автоматично
dataset.commit.note	Опис версії набору	
dataset.commit.HEAD	Ознака того, чи є ця версія актуальною	
dimension.*	Опис напрямків куба даних	
dimension.country.*	Приклад опису напрямку "country"	Заповнюється для кожного напрямку
dimension.country.label	Назва напрямку	
dimension.country.scale	Тип шкали	nominal, ordinal, interval
dimension.country.role	Роль	geo (є геокодування), time (час), metric (вимірювання)
dimension.country.unit	Одиниця вимірювання	Для інтервальної шкали
layout.*	Опис способу розбору даних	Для кожного типу набору використовується власна структура (нижче приклад для багатовимірного куба)
layout.sheet	Назва аркуша з даними	
layout.value	Назва стовпця з значеннями	
layout.country.id	Назва стовпця з ідентифікаторами об'єктів	Заповнюється для кожного напрямку
layout.country.label	Назва стовпця з назвами об'єктів	Заповнюється для кожного напрямку

У випадку необхідності розширення метаданих можна використати тлумачний словник, елементи якого ставлять набір значень у відповідність до ключа. При цьому значення ключів можна використовувати як значення метаданих. Наприклад, для *dataset.source* = “#WB” у словнику можна задати структуру (дивись табл. 4.2), яка описує назву, короткий опис, картинку та посилання на сайт порталу даних Світового Банку.

Таблиця 4.2.

Структура словника метаданих

key	value.label	value.note	value.icon	value.url
#WB	#WB_NAME	#WB_NOTE	https://innovation.internews.org/sites/default/files/styles/ncil_large/public/content-images/world-bank-logos-282.jpg	http://data.worldbank.org/

Аналогічно описується локалізація рядків (приклад локалізації для рядків з попередньої таблиці наведений в таблиці 4.3)

Таблиця 4.3

Структура локалізації

key	value.en	value.ua
#WB_NAME	World Bank	Світовий Банк
#WB_NOTE	World Bank Open Data: free and open access to data about development in countries around the globe.	World Bank Open Data: вільний і відкритий доступ до даних про розвиток в країнах по всьому світу.

Приклад набору даних в форматі JSON, який використовується для обміну даними та збереження наборів наведено нижче.

```
{
  "dataset": {
    "id": "3a37ea80-93a8-11e5-b62f-dfcea48fc8d9",
    "status": "public",
    "commit": {
      "id": "579b1b6b13abd85813a20173",
      "author": "Andrey Boldak",
      "note": "Add UA flag",
      "HEAD": true,
      "createdAt": "2016-07-29T09:01:31.070Z"
    },
    "label": "#GDP_LABEL",
    "note": "#GDP_NOTE",
    "source": "#WB",
    "topics": [
      "#GDP",
      "#WDI/#EC/#GDP",
      "#EC/#GDP",
      "#WDI/#GDP",
      "That's all folks"
    ]
  },
  "locale": [
    "#en",
    "#ua"
  ],
  "dimension": {
    "country": {
      "label": "#DIM_COUNTRY",
      "role": "geo",
    }
  }
}
```

```

"values": [
  {
    "id": "LBY",
    "label": "Libya"
  },
  {
    "id": "ABW",
    "label": "Aruba"
  },
  {
    "id": "AGO",
    "label": "Angola"
  }
],
"concept": {
  "label": "#DIM_INDICATOR",
  "role": "metric",
  "values": [
    {
      "id": "GDP",
      "label": "GDP per capita (current US$)"
    }
  ]
},
"year": {
  "label": "#DIM_YEAR",
  "role": "time",
  "values": [
    {
      "id": "2010",
      "label": "2010"
    }
  ],
  {
    "id": "2011",
    "label": "2011"
  },
  {
    "id": "2012",
    "label": "2012"
  },
  {
    "id": "2013",
    "label": "2013"
  }
]
},
"layout": {
  "sheet": "data",
  "value": "Value",
  "country": {
    "id": "Country Code",
    "label": "Country Name"
  },
  "concept": {
    "id": "Indicator Code",
    "label": "Indicator Name"
  },
  "year": {
    "id": "Year",
    "label": "Year"
  }
}
}

```

Для опису структурованих послідовностей подій введено тип наборів даних «Timeline», формат метаданих якого описано в таблиці 4.4.

Таблиця 4.4.

Структура метаданих для часових послідовностей подій

Підструктура метаданих	Призначення	Примітка
dataset.*	Опис набору даних	
dataset.label	Назва набору	
dataset.note	Опис набору	
dataset.id	Універсальний унікальний ідентифікатор набору	Попередньо генерується підсистемою довготривалого зберігання
dataset.type	Тип набору даних	timeline
layout.*	Опис способу розбору даних	
layout.dateFormat	Формат дати	DD – день (01, 31), ММ – місяць (01, 12), YYYY – рік (2016)

Продовження Таблиці 4.4

Підструктура метаданих	Призначення	Примітка
layout.sheet	Аркуш з даними	
layout.id	Стовпчик "Ідентифікатор події"	
layout.start	Стовпчик "Початок події"	Дата у відповідному форматі
layout.end	Стовпчик "Кінець події"	Дата у відповідному форматі
layout.category	Стовпчик "Категорія події"	Події можна класифікувати за допомогою категорій
layout.hierarchy	Стовпчик "Ієрархія події"	Події класифікуються через їх ієрархію. Клас події визначає шлях до неї
layout.type	Стовпчик "Тип події"	короткочасна подія, потік подій, довготривалий процес
layout.context	Стовпчик "Опис події"	Як значення використовуються ключі словника

Приклад метаданих в форматі JSON, які описують послідовності подій, наведено нижче:

```

"dataset": {
  "label": "Breaks",
  "note": "Breaks minimization",
  "type": "timeline"
},
"layout": {
  "dateFormat": "DD MM YYYY",
  "sheet": "data",
  "id": "id",
  "start": "start",
  "end": "end",
  "category": "category",
  "hierarchy": "id",
  "type": "type",
  "context": "id"
}

```

Фрагмент аркуша даних наведено в таблиці 4.5.

Таблиця 4.5

Приклад даних для часових послідовностей подій

id	category	type	start	end
t1	c1	instant	16 01 2014	
t2	c1	instant	28 01 2014	
t3	c1	process	18 02 2014	22 02 2014
t4	c1	instant	27 02 2014	
t5	c2	instant	16 03 2014	
t6	c2	process	05 04 2014	15 04 2014

4.1.2. Підсистема опрацювання даних

Ця підсистема забезпечує ефективне функціонування мікросервісів, надаючи їм контейнер для виконання з необхідним рівнем ізолюваності обчислювального процесу. Взаємодія серверів обробки даних між собою та з іншими складовими платформи здійснюється через REST-інтерфейс. Кеш запитів на виконання окремих мікросервісів забезпечує зниження навантаження на сервер за рахунок блокування повторного виконання запитів. Контейнер мікросервісів забезпечує їх виконання в дочірніх обчислювальних процесах. Взаємодія між контейнером та мікросервісом здійснюється завдяки обміну повідомленнями. Модуль управління мікросервісами здійснює загальне управління (orchestration) мікросервісами власного та інших серверів обробки даних під час реалізації сценаріїв обробки даних, заданих в паралельно-послідовній формі.

4.1.3. Загальний реєстр ресурсів

Ця підсистема реалізує покладені на неї функції реєстрування мікросервісів, сховища мікросервісів, що можуть переміщуватися та допускають розгортання багатьох екземплярів, динамічної реконфігурації серверів оброблення даних за рахунок перерозподілу таких мікросервісів. Також реєстр виконує функції диспетчеризації запитів на виконання сценаріїв, перенаправляючи їх в модулі управління мікросервісами.

4.1.4. Підсистема швидкого розроблення і підтримки прикладних тематичних застосунків

Ця підсистема забезпечує для кінцевого користувача інтерфейс для створення застосунків на основі використання віджетів, що можуть взаємодіяти з серверами опрацювання даними. Практичне вирішення цього завдання ґрунтується на вдосконаленні розроблено в СІД України прототипу, вихідні коди якого доступні в депозитарії GitHub (<https://github.com/kpi-wdc/wdc>), призначеного для «швидкої розробки» прикладних застосунків

(технологія Single Page Application) на основі використання віджетів, пов'язаних з інтелектуальним опрацюванням та візуалізацією даних.

Специфіка підтримки міждисциплінарних досліджень у ССД полягає у тому, щоб надати науковцям, які використовують дані різних СЦД, відповідні інструменти. Ці можливості реалізує платформа. Але, якщо користувач має певні навички використання ІТ, йому можна надати інструменти, які значно пришвидшують отримання результатів, при цьому створюючи умови для створення і повторного використання (не тільки цим користувачем, але й іншими) нового інформаційного ресурсу у вигляді тематичного застосунку. Націленість платформи на підтримку міждисциплінарних досліджень підказує перший і досить поширений тип таких застосунків – швидке приготування аналітичних звітів різної форми з різних джерел. Тому підсистема швидкого розроблення і підтримки прикладних тематичних застосунків створена для того, щоб надати користувачеві набір інструментів, необхідний для створення інтерактивних багатосторінкових звітів з результатами досліджень на основі аналітичного опрацювання даних.

У відповідності до цієї мети в підсистемі реалізовані наступні вимоги:

- online-розробка односторінкових застосунків;
- хостинг односторінкових застосунків;
- персональні тематичні панелі на основі використання віджетів;
- взаємодія віджетів;
- гнучка схема навігації;
- гнучка (незалежна від пристроїв) компоновка сторінок;
- стилі застосунків, повна інтернаціоналізація застосунків;
- управління ресурсами застосунків;
- підтримка взаємодії користувачів;
- авторизація доступу до ресурсів.

Односторінкові web-застосунки використовують віджети – налагоджувані елементи інтерфейсу користувача, призначені для вирішення специфічної задачі. Якщо вибрати деякий бізнес-процес, наприклад процес

«Управління даними», тоді набір тематичних віджетів дозволить його автоматизувати. При цьому схема компоновки тематичної панелі може бути довільною.

У складі платформи реалізовані набори інструментів, які призначені для:

- управління застосунками;
- управління даними: підтримка розширюваного набору метаданих, повна інтернаціоналізація наборів даних, контроль версій даних, імпорту/експорту даних;
- інтелектуального опрацювання та візуалізації даних: маніпулювання багатовимірними кубами даних, запити (проекції, вивірки, злиття), інтерактивні графіки, багатoshарові тематичні мапи, таблиці, вбудовані інструменти опрацювання та аналізу даних;
- аналізу часових послідовностей подій.

Приклад такого застосунку наведено на рис. 4.2.

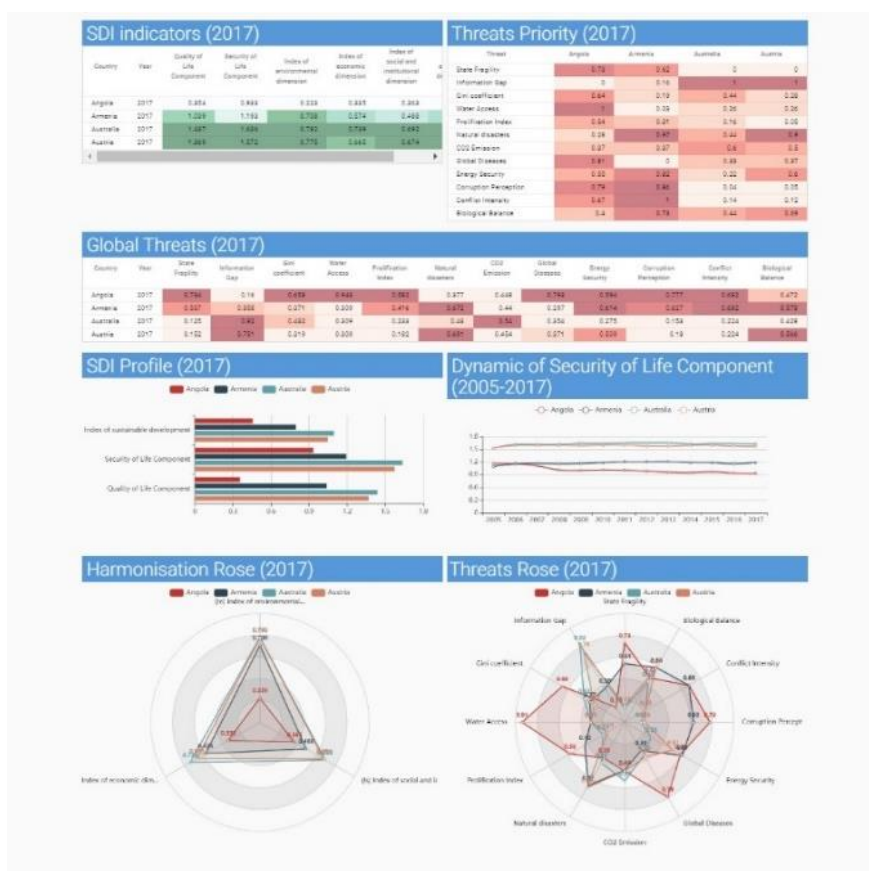


Рис. 4.2. Приклад візуалізації даних за допомогою засобів платформи підтримки міждисциплінарних досліджень

4.2. Реалізація платформи підтримки міждисциплінарних досліджень

4.2.1. Виділення і поділ структурних компонентів. Архітектура розробленої платформи підказує важливі ідеї, корисні з точки зору її реалізації. По-перше, було б вигідно класифікувати її компоненти на основі певних ознак. Тоді компоненти одного класу можна було б реалізувати швидше і з меншими витратами. Один з варіантів виділення компонентів наведений на рис. 4.3.

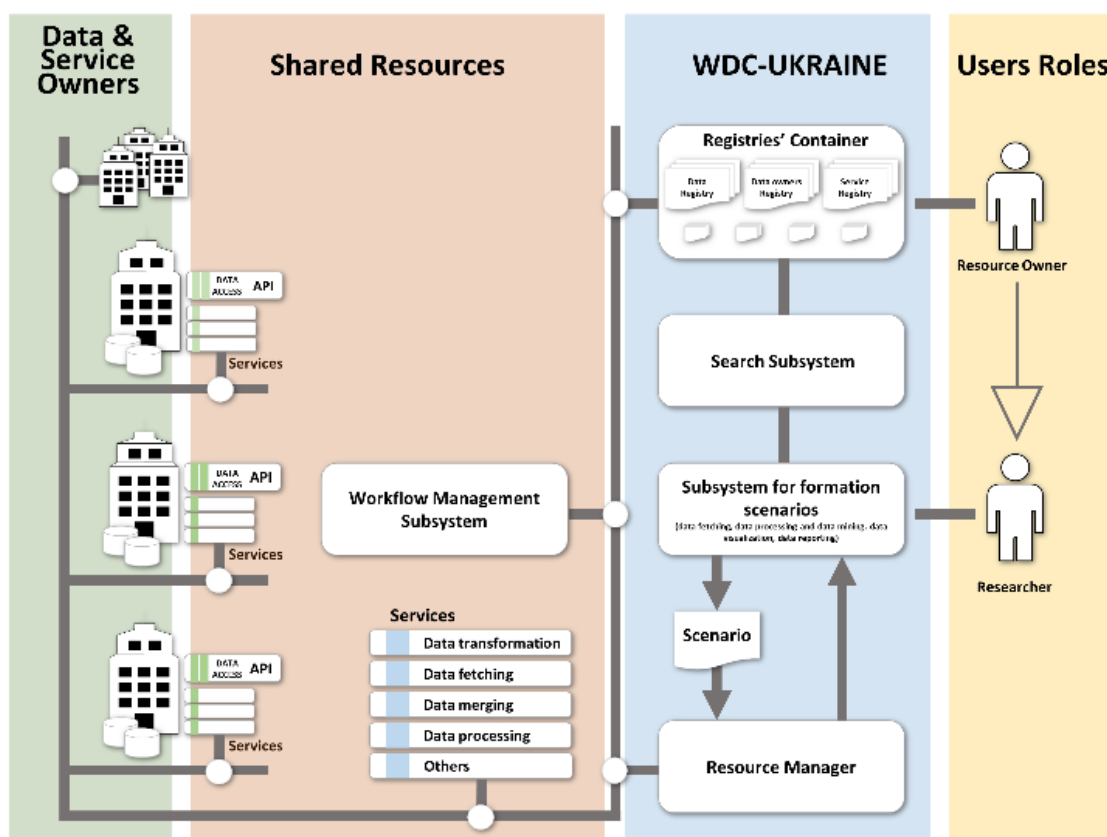


Рис. 4.3. Структурна схема компонентів платформи

Компоненти архітектури платформи (high level architecture) поділені на 4 сфери:

- дані і власники даних і сервісів;
- ресурси, які поділяються спільнотою ССД – сервіси даних, сервіси опрацювання даних (перетворення, пошуку і вибирання; злиття; оброблення та інші) та засоби управління потоками даних;

- компоненти СЦД-Україна – контейнери ресурсів, менеджер ресурсів, компоненти пошуку і формування сценаріїв;
- ролі учасників спільноти (власники, дослідники).

Для забезпечення організації міждисциплінарного дослідження створювана платформа повинна надавати щонайменше інструменти:

- реєстрації партнерів в системі як суб'єктів – джерел даних;
- формування та управління власними предметно-орієнтованими розділами (для даних) в рамках загальної онтології даних;
- перетворення даних з одного формату в інший;
- надання прив'язки даних до конкретних науково-дослідних проектів;
- публікації даних інструментів, що підтримують призначення прав доступу, що контролюється власником даних.

Крім того, для забезпечення повторного використання даних повинні надаватися базова онтологія пошукової підсистеми та інструменти для вибору даних та об'єднання на основі результатів пошуку.

З огляду на специфіку накопичених даних, ця платформа має надавати користувачам інструменти розгортання та реалізації методів аналізу даних. Останні можна розділити на загальні та спеціальні. Перші змінюються після їх імплементації і запуску. Але рейтингування, відстеження помилок, зберігання історії використання в різних завданнях виконуються для кожного методу.

Це обумовлює формування для міждисциплінарних досліджень єдиної методології аналізу даних. Важливою рисою платформи також є можливість формування і виконання складних послідовностей обробки даних – сценаріїв.

4.2.2. Технічні та технологічні аспекти підтримки міждисциплінарних досліджень

Інформаційне середовище СЦД-Україна базується на платформі підтримки міждисциплінарних наукових досліджень і становить собою розподілену систему інформаційно-аналітичних систем, призначену для підтримки міждисциплінарних наукових досліджень. Нижній рівень ієрархії системи забезпечує стандартизацію процесів обслуговування даних, що є однією з обов'язкових умов успішної сертифікації центрів у ССД. Ці процеси утворюють безперервний життєвий цикл із наступними фазами: створення, оброблення, аналіз даних, їх збереження, публікація й повторне використання [136-138]. Прийнята ССД концепція інтеграції джерел даних і сервісів – Global System of Data Systems вимагає представляти інформаційні ресурси всієї сукупності інтегрованих джерел і сервісів як нове єдине джерело [139-141].

Для розв'язання цієї задачі у роботі розроблено платформу підтримки міждисциплінарних наукових досліджень, яка включає компонент інтеграції даних і сервісів на семантичному рівні з використанням мови OWL. Він забезпечує підтримку єдиного подання даних з врахуванням їх семантичних властивостей у контексті єдиної онтології предметної області [103]. У цій системі джерела даних каталогізуються за допомогою онтології Global Change Master Directory (GCMD) Science Keywords [142].

Інтеграція даних і сервісів реалізована на основі агентно-орієнтованого підходу, який у рамках сервіс-орієнтованої архітектури забезпечує інтеграцію даних і сервісів, організацію їх взаємодії через агентів [103]. Агенти є спеціалізованими мігруючими програмними компонентами, які мають унікальні Owl-онтології і файли XML, що визначають параметри підключення до джерела, правила проектування словника на джерело і параметри функціонування [53].

Взаємодія з розробленою платформою відбувається через портал СЦД-Україна: <http://wdc.org.ua>. На рис. 4.4 представлено зовнішній вигляд Web-інтерфейсу користувача, який забезпечує організацію віртуального простору

джерел даних і сервісів: реєстрація джерел даних і сервісів із прив'язкою до єдиної GCMD-онтології, моніторинг стану джерел даних і сервісів, одержання довідкової метаінформації про них; пошук і інтеграцію даних: формування запитів шляхом завдання параметрів фільтрів або безпосереднього введення SPARQL-запитів, обмеження зони пошуку за джерелами даних, приведення результатів пошуку до вигляду, обумовленого користувачем (підтримуються формати XML, JPEG, CSV, HTML, різні види графіків і таблиць). Також система містить у собі сервіси для аналітичної обробки даних (факторний аналіз, кластерний аналіз, кореляційний аналіз та інші).

На рис. 4.5 зображено технологічний процес інтелектуального опрацювання даних, який підтримується в СЦД-Україна [11, 13]. Цей процес можна розділити на стадії попередньої обробки, аналізу та пост-обробки даних, на кожній з яких вирішуються специфічні задачі з використанням методів багатовимірного статистичного аналізу, статистично-експертних методів, тощо.

Оскільки одні й ті самі методи використовуються для вирішення різних задач, а одна задача може вирішуватися різними методами, виділені програмні засоби, які реалізують статистичні та експертні методи незалежно від контексту їх застосування для вирішення конкретних задач оброблення даних.

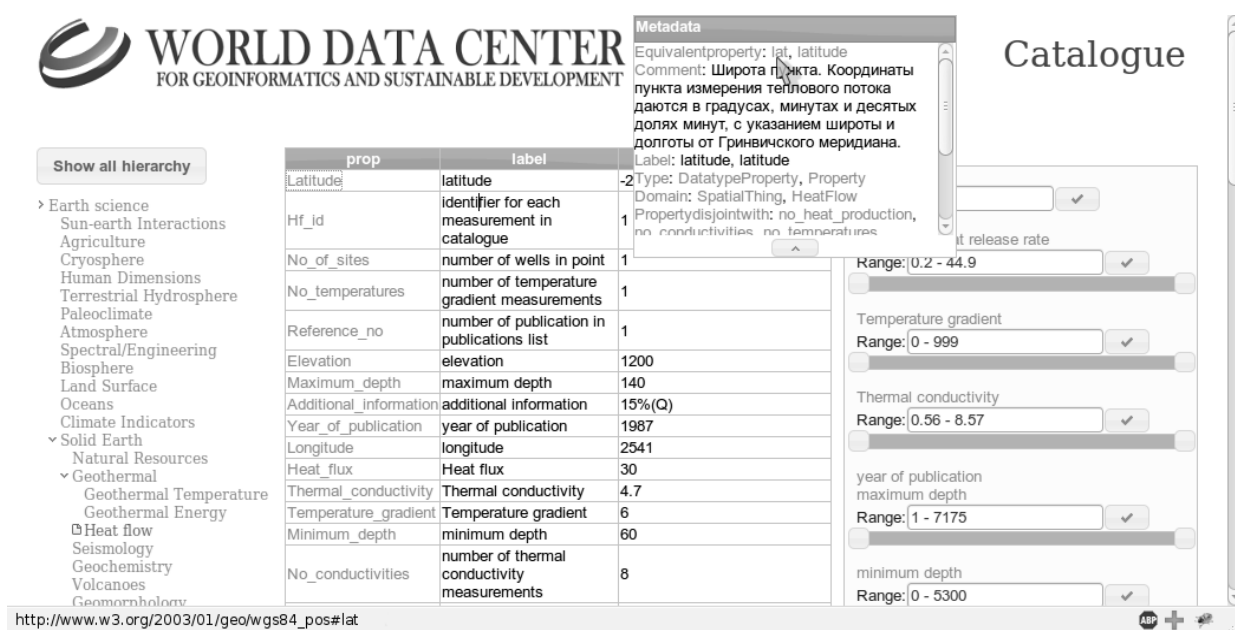


Рис.4.4. Web-інтерфейс користувача для пошуку та інтеграції даних.

		Процес обробки даних			Підготовка звітів та візуалізація даних	
		Попередня обробка даних		Аналіз даних		
Статистичні та експертно-статистичні методи	Основні статистики	Оцінка помилок	Визначення відсутніх значень	Нормування даних	Визначення інтервалів для візуалізації даних	
	Критерії узгодженості	Аналіз надійності та достовірності даних		Фільтрація даних		
	Аналіз головних компонентів			Розробка показників	Аналіз залежностей та редукція моделі	Редукція параметричного простору та вибір проєкцій
	Факторний аналіз			Розробка показників	Аналіз залежностей	
	Байєсівські мережі довіри			Дослідження причинно-наслідкових зв'язків		Формальна інтерпретація результатів моделювання причинно-наслідкових зв'язків
	Ентропійний аналіз			Розробка показників	Проектування діаграм впливу	
	Кореляційний аналіз		Визначення відсутніх значень	Розробка показників	Аналіз залежностей та проектування діаграм впливу	
	Кластеризація		Визначення відсутніх значень		Формальна класифікація	
	Лінійна та нелінійна регресія		Визначення відсутніх значень		Оцінювання параметрів моделей	

Рис.4.5. Процес інтелектуального опрацювання даних.

З них сформовані бібліотеки програм, використовуваних для реалізації задач оброблення даних в межах загального процесу, зображеного на рис. 4.5. За наявності достатнього спектру засобів можна говорити про реалізацію концепції «швидкого» розроблення програм для вирішення різноманітних задач інтелектуального оброблення даних. Для цього у роботі розроблено проблемно-орієнтовану графічну мову аналітичного оброблення даних, використання якої не потребує спеціальних знань та навичок програмування. Експерти можуть розробляти нові та модифікувати існуючі сценарії без залучення програмістів [7]. Досвід показав, що графічне подання процесу аналітичного опрацювання даних, його автоматична трансформація до вигляду, придатного до виконання в інструментальному середовищі, наявність інтерактивних засобів налагодження дозволяють знижують трудомісткість розробки процедур попереднього оброблення великих обсягів даних, оцінювання, формування аналітичних звітів.

4.2.3. Інтерфейс користувача на прикладі підсистеми створення застосунків. Користувач авторизується за допомогою профілю Google на стартовій сторінці застосунку «Apps» і натискає кнопку «Створити застосунок» (рис. 4.6).

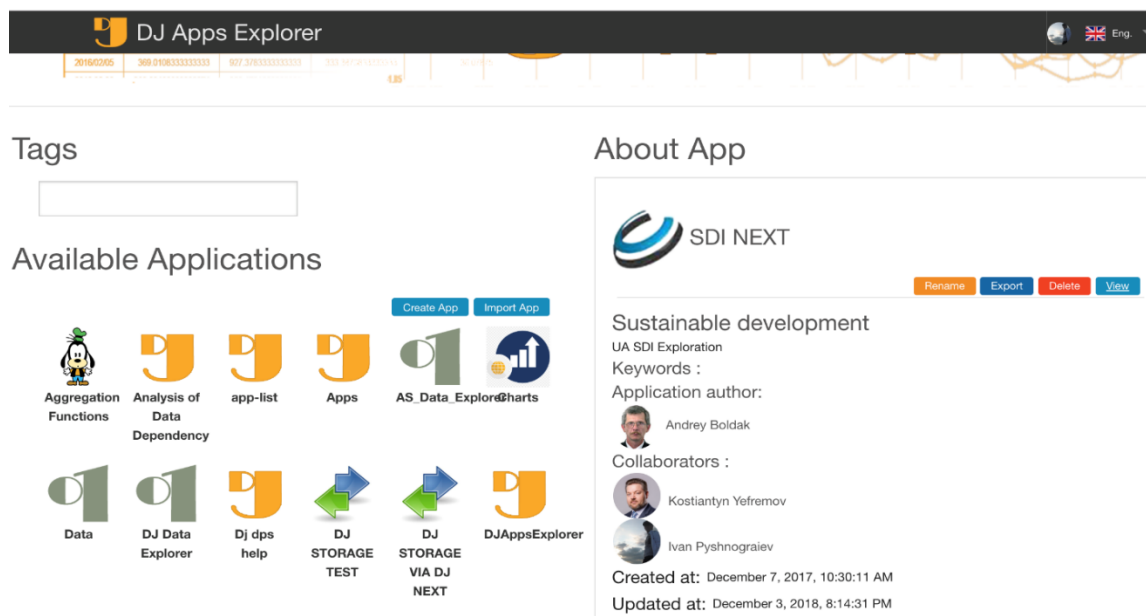


Рис. 4.6. Стартова сторінка застосунку «Apps»

Користувач може задати назву застосунку, вибрати доступний шаблон (рис. 4.7). Система створить застосунок і відобразить його на панелі (рис.4.8).

До налаштування застосунку (який може знаходитись в стані «Дизайн» або «Презентація») перехід здійснюється натисканням кнопки «Перегляд» (рис.4.9).

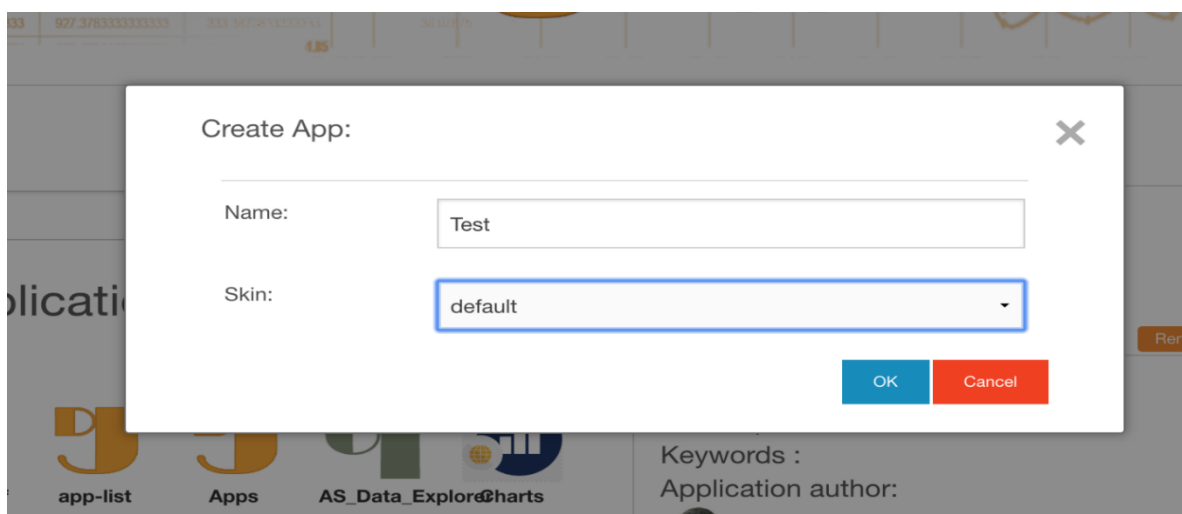


Рис. 4.7. Діалогове вікно створення нового застосунку

Перемкнути режим можна кнопками панелі розробника (рис.4.9), а налаштувати застосунок – кнопками панелі «Налаштування» (рис. 4.10)

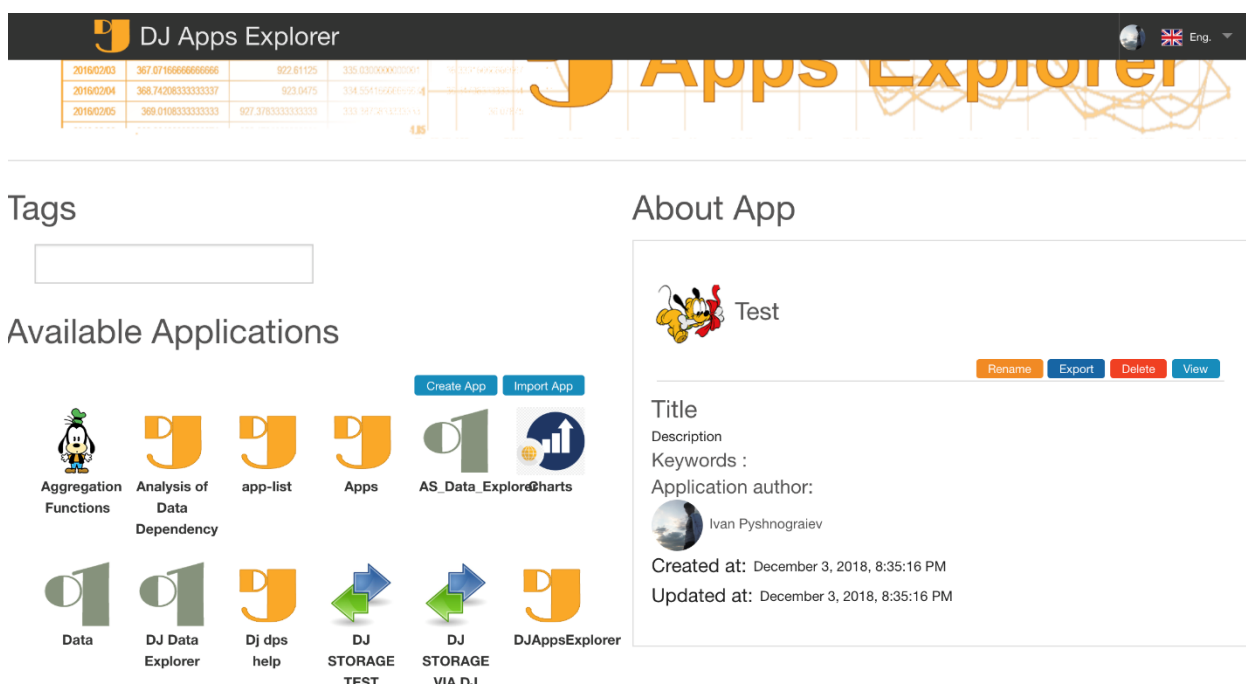


Рис. 4.8. Відображення нового застосунку на стартовій сторінці «Apps»

По налаштуваннях «Apps» використовує метадані застосунку для пошуку та відображення короткої та докладної інформації про нього (рис.4.11).

Панель розробника надає можливість перейти до налаштування сторінок, завантажених ресурсів та інтернаціоналізації застосунку (рис. 4.9).

Під час створення нової сторінки користувач може задати повну та коротку назви сторінки та вибрати один з доступних шаблонів (рис. 4.12).

Управління сторінками здійснюється після натискання на кнопку «Сторінки» (рис. 4.13).

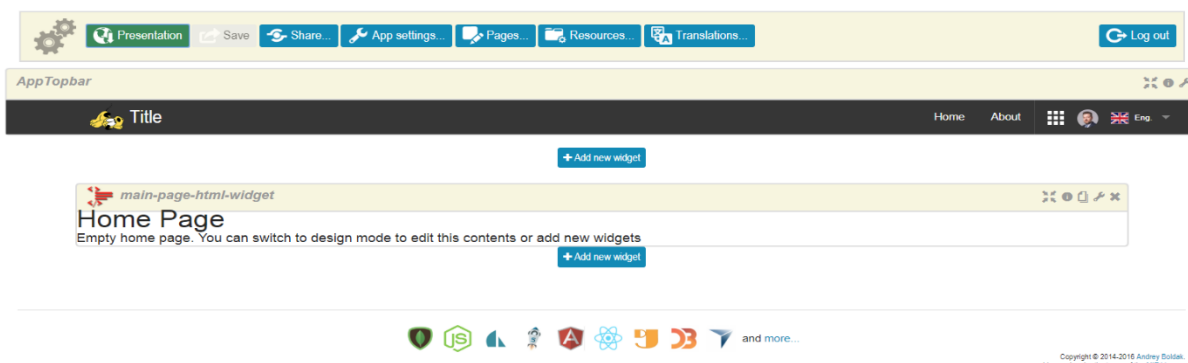


Рис. 4.9. Панель розробника

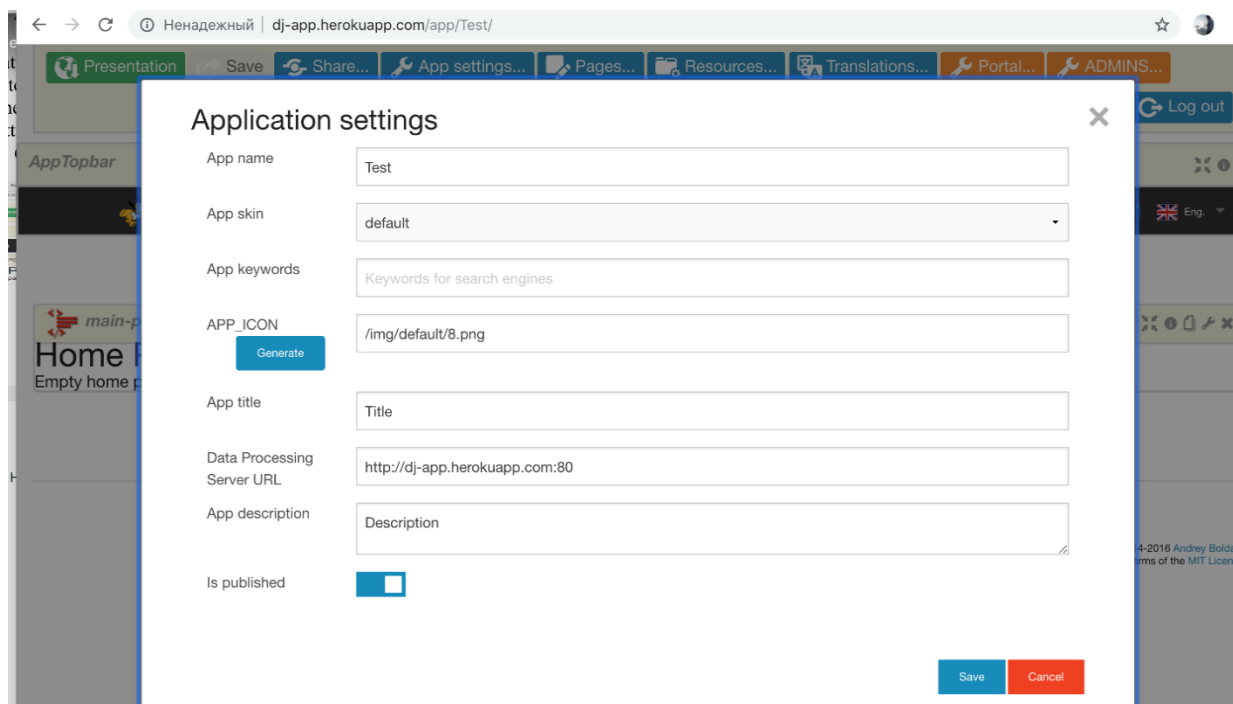


Рис. 4.10. Меню «Загальні налаштування застосунка»

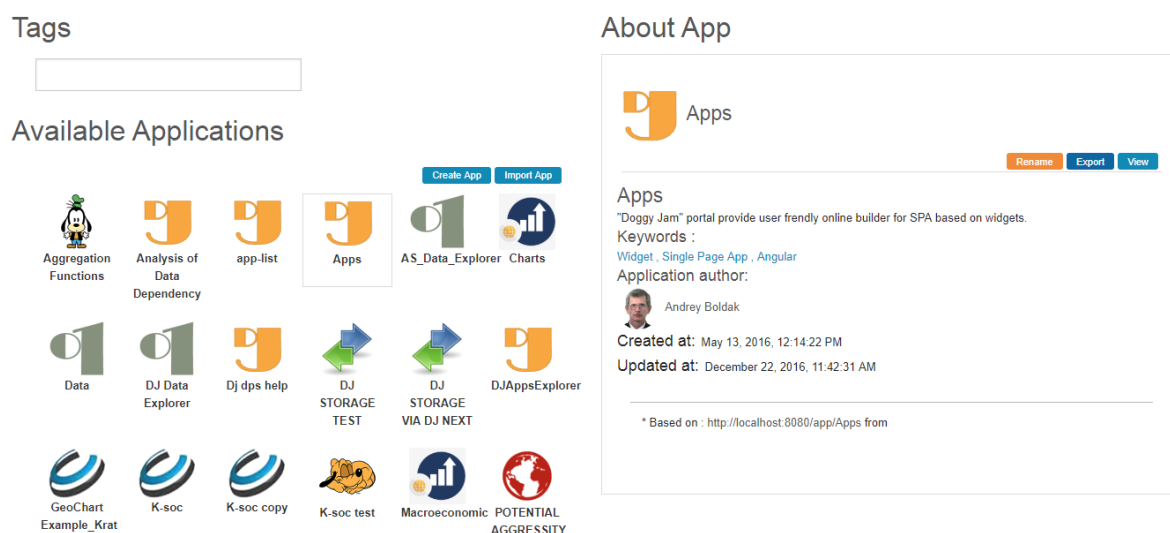


Рис. 4.11. Відображення інформації про застосунок на стартовій сторінці застосунку «Apps»

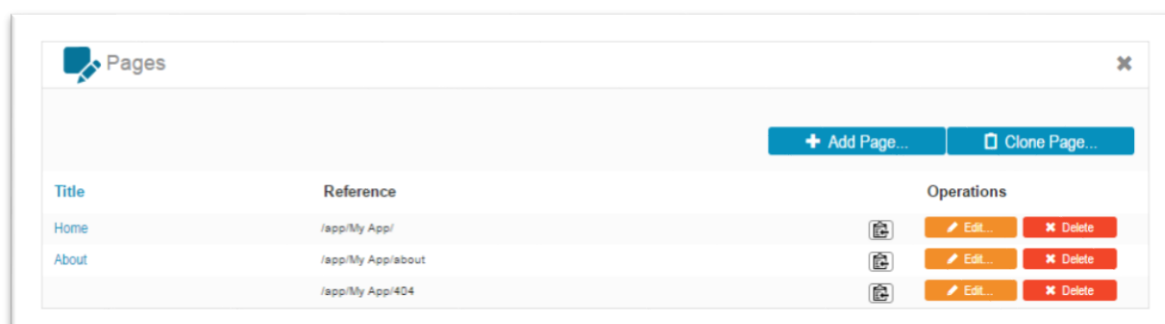


Рис. 4.12. Діалогове вікно створення нової сторінки

Після налаштування система створює сторінку з вибраним шаблоном, який містить контейнери для віджетів, в кожен з яких після натискання відповідної кнопки можна вставити віджети (рис. 4.14). Панель з доступними віджетами наведено на рис. 4.15.

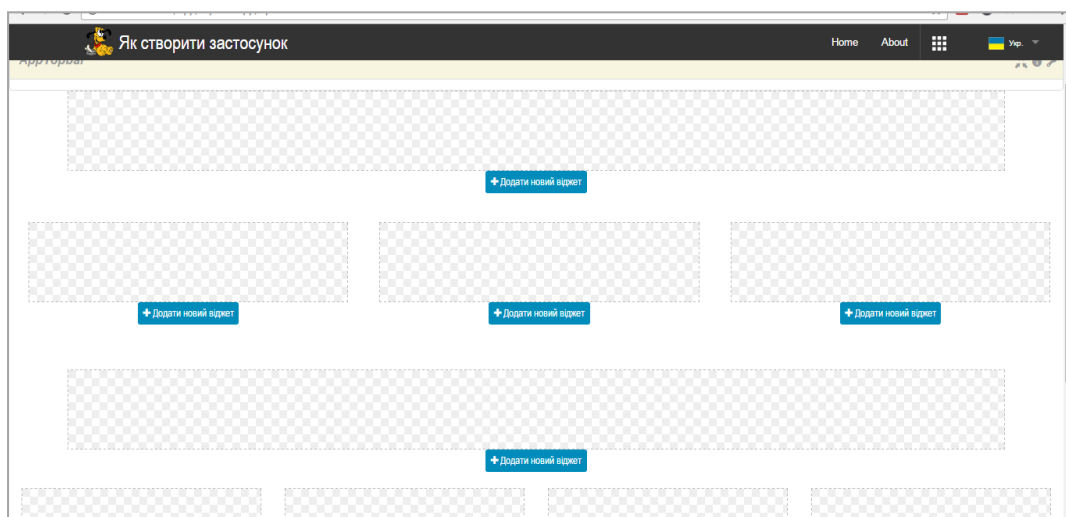


Рис. 4.13. Меню управління сторінками

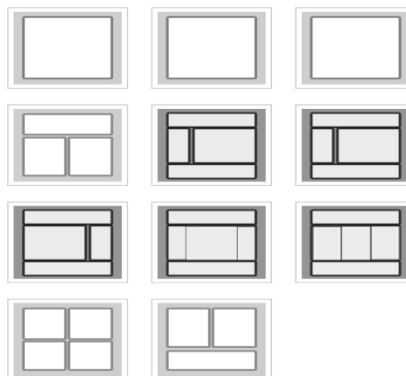
Add a new page

Short title

Unique link (leave empty for home page)

Template

WIDGET.PAGE-LIST.SE



Add

Cancel

Рис. 4.14. Нова сторінка з контейнерами для віджетів

Після вибору віджета користувач може змінити його налаштування (рис. 4. 16). Після їх збереження віджет відображується на сторінці (рис. 4.17).

Приклад застосунку, створеного для побудови стратегій та сценаріїв сталого розвитку України, наведено на порталі СЦД-Україна. Отримати доступ можна за адресою <http://open.wdc.org.ua/>. Застосунок складається з низки інтерактивних тематичних панелей (dashboard) та дозволяє в зручній для

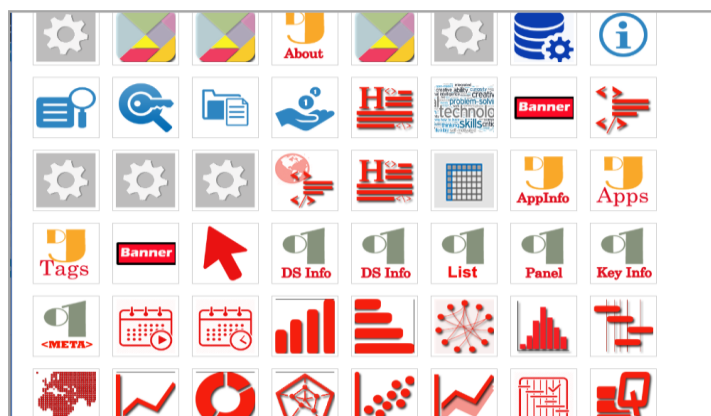


Рис. 4.15. Панель віджетів

Scatter Chart Settings Wizard

Steps: **0. Widget ID** 1. Chart Decoration

Edit widget ID if needed

Widget ID:

pauraxgrvf

Next Cancel

Рис. 4.16. Меню налаштування віджету

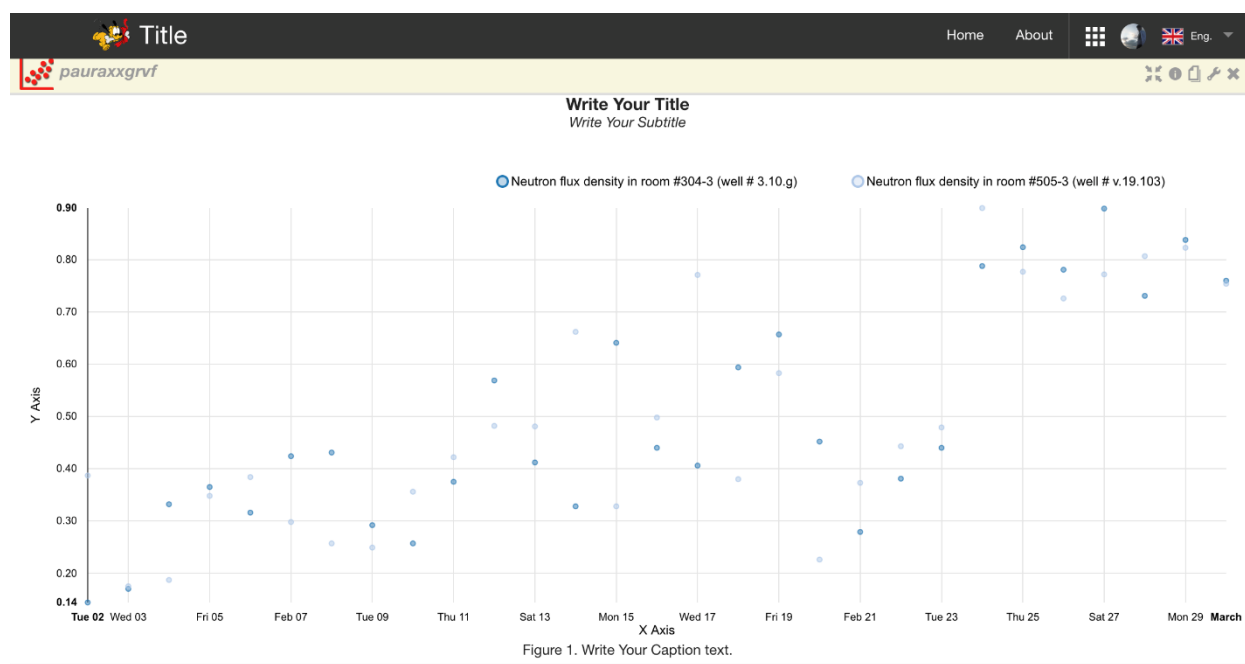


Рис.4.17. Приклад відображення віджету на сторінці застосунку

користувача формі моделювати та здійснити порівняльний аналіз показників сталого розвитку України в двох розрізах:

1) глобальному, коли індикатори сталого розвитку України порівнюються з відповідними індикаторами інших країн Європи і світу;

2) регіональному, коли проводиться моделювання і порівняльний аналіз сталого розвитку регіонів (областей) України.

Результати цього моделювання і порівняльного аналізу використовуються для визначення наближення (чи віддалення) соціально-економічного розвитку України до бажаних сценаріїв «Оптимістичного» (до 2020 року) і «Збалансований розвиток» (до 2030 року) [8].

4.3. Застосування платформи для міждисциплінарних досліджень на основі моделей і методів інтелектуального аналізу даних із різних джерел

У підрозділі розглянуто модель кількісного оцінювання процесів сталого розвитку в контексті якості та безпеки життя людей, наведено результати моделювання для країн світу та адміністративних регіонів України. Представлено результати моделювання процесів сталого розвитку країн світу та адміністративних регіонів України, отриманих з використанням засобів платформи підтримки міждисциплінарних досліджень.

Метрики (індикатори й індекси) для вимірювання сталого розвитку сприяють переведенню знань у фізичних і соціальних науках у керовані інформаційні блоки, які оснащують процес ухвалення управлінських рішень. Вони дозволяють виміряти й оцінити прогрес у досягненні цілей сталого розвитку і забезпечити інформування та своєчасне попередження з метою запобігання критичного стану й збитків економіки, проблем у соціальній і природоохоронній сферах. Важливими є задачі створення моделей оцінювання сталого розвитку, як орієнтованих на глобальний контекст, так і адаптованих до національних пріоритетів, національної та регіональної екологічної, економічної і соціальної політик. Основою для розв'язання таких міждисциплінарних задач є великі обсяги даних за багатьма геофізичними і гуманітарними дисциплінами та видами спостережень, а також організація спільної дослідницької діяльності великої кількості наукових організацій, колективів та окремих експертів з різних наукових напрямків.

Головними особливостями опрацювання даних із різних джерел у складі розробленої платформи є, з одного боку, вибір і розроблення відповідних моделей і методів інтелектуального аналізу даних, з іншого боку, засобів приведення даних із різних джерел до єдиної шкали.

Також у підрозділі представлено підхід до оцінювання ефективності суспільних перетворень як міри неузгодженості між діями влади і очікуваннями суспільства та синергії (соціальної активності) людей, заснований на формалізованому узгодженні результатів, отриманих методом експертних оцінок і методами сентимент-аналізу та інтелектуального аналізу неструктурованих даних великих обсягів, зокрема текстових повідомлень з відкритих онлайн-джерел й соціальних мереж [17]. Ці методи реалізовано у вигляді комплексу web-сервісів та застосунків засобами розробленої в рамках дисертаційної роботи платформи підтримки міждисциплінарних досліджень.

За даними провідних міжнародних дослідницьких та консалтингових компаній International Data Corporation (IDC), International Data Group (IDG) неструктуровані дані різних типів у своїх обсягах досягають 80% – 90% від загального обсягу цифрових даних, що зберігаються в усьому світі [143, 144]. Застосування методів оброблення текстових даних природньої мови [145-147] дозволяє отримати атрибутивний опис цих текстових даних у вигляді структурованих інформаційних об'єктів, які відповідають семантиці виявлених фактів. В свою чергу до такого атрибутивного опису під час аналізу управлінських ситуацій можна застосовувати статистичні методи, методи інтелектуального аналізу даних, машинного навчання, пропоновані в роботі методи системного узгодження даних тощо. Такий підхід до оброблення та використання неструктурованих даних дозволяє отримувати більш повні та адекватні моделі управлінських ситуацій в сфері економіки, суспільних відносин, національної безпеки та оборони в умовах багатокритеріальності розв'язуваних задач, багатофакторності ризиків, неповноти і невизначеності даних, наявності великої кількості обмежень, та використовувати ці моделі для підготовки ефективних стратегічних управлінських рішень [17, 148-151].

Продemonстровано застосування розробленого в роботі інструментарію інтелектуального аналізу та системного узгодження міждисциплінарних даних на прикладах моделювання процесів сталого розвитку країн світу та адміністративних регіонів України (п. 4.3.1) і кількісного оцінювання ставлення населення України до проведення масової вакцинації проти COVID-19 (п. 4.3.2).

4.3.1. Моделі і методи системного узгодження даних і моделювання процесів сталого розвитку України

Характерним прикладом використання розроблених в дисертаційному дослідженні принципів системного узгодження даних різної природи є завдання моделювання сталого розвитку територій (країн, регіонів, адміністративно-територіальних одиниць тощо) в контексті якості та безпеки життя людей та оцінювання ступеня впливу сукупності загроз на їх розвиток.

Відповідно до метрики, запропонованої в роботах [12, 13], процес сталого розвитку характеризується двома основними складовими: безпекою (C_{sl}) і якістю (C_{ql}) життя людей, а узагальнена міра сталого розвитку визначається за допомогою кватеріона:

$$\{Q\} = jw_{sl}C_{sl} + w_{ql}\|\overrightarrow{C_{ql}}\|. \quad (4.1)$$

Кватеріон $\{Q\}$ містить уявну зважену скалярну частину $jw_{sl}C_{sl}$, яка описує безпеку життя людей і зважену дійсну векторну частину, яка описує якість життя людей у просторі трьох вимірів: економічного (I_e), екологічного (I_{ec}) і соціально-інституціонального (I_s). При цьому j набуває значення дійсної одиниці для нормального, регулярного стану розвитку суспільства за $C_{sl} > 0$, і значення уявної одиниці, коли суспільство переходить у стан конфлікту ($C_{sl} = 0$). Згідно з результатами аналізу стану країн світу Heidelberg Institute for International Conflict Research визначив такі рівні конфліктів [152]:

- 0 (Without crisis) – у країні відсутні конфлікти;
- 1 (Dispute) – політичний конфлікт без людських втрат;

- 2 (Non-violent crisis) – політичний конфлікт, в якому одна з сторін погрожувала розправою;
- 3 (Violent crisis) – слабоінтенсивний конфлікт з невеликими людськими втратами, в якому використовується тільки легка зброя;
- 4 (Limited war) – інтенсивний конфлікт з середніми людськими втратами, в якому або використовується тяжка зброя або задіяно значну кількість учасників;
- 5 (War) – інтенсивний конфлікт з великими людськими втратами, в якому використовується тяжка зброя та задіяно значну кількість учасників.

Тому j визначається наступним чином:

$$j = \begin{cases} 1, & \text{для 1 – 4 рівней конфлікту;} \\ \sqrt{-1}, & \text{для 5 рівня конфлікту.} \end{cases}$$

Вагові коефіцієнти w_{sl} і w_{ql} в формулі (4.1) використано з метою вирівнювання масштабів компонент безпеки і якості життя (у випадку оцінювання сталого розвитку країн світу – $w_{sl} = 1/\sqrt[3]{11}$, $w_{ql} = 1/\sqrt{3}$).

Для кожної території евклідова норма радіус вектора якості життя людей (\vec{C}_{ql}) подається у такому вигляді:

$$\|\vec{C}_{ql}\| = \sqrt{I_{ec}^2 + I_e^2 + I_s^2} \quad (4.2)$$

Кут відхилення α радіус вектора \vec{C}_{ql} від ідеального вектора (1,1,1) визначають через значення вимірів I_{ec}, I_e, I_s у такий спосіб:

$$\alpha = \arccos \frac{I_{ec} + I_e + I_s}{\sqrt{3} \cdot \sqrt{I_{ec}^2 + I_e^2 + I_s^2}}, \quad (4.3)$$

$$0 \leq \alpha \leq \arccos \frac{1}{\sqrt{3}}.$$

Отже, норма радіус вектора \vec{C}_{ql} характеризує якість життя людей, а просторове положення вектора \vec{C}_{ql} у системі координат (I_{ec}, I_e, I_s) характеризує міру «гармонійності» сталого розвитку. Рівновіддаленість

вектора \vec{C}_{ql} від кожної з зазначених координат буде відповідати найбільшій гармонійності сталого розвитку, а наближення цього вектора до однієї з координат буде вказувати на пріоритетний розвиток за відповідним виміром і нехтування двома іншими.

Величина $G=I-\alpha$ називається ступенем гармонізації сталого розвитку. Вона зростатиме у разі наближення G до 1 і зменшуватиметься у разі наближення G до 0.

Таким чином, компонента якості життя – інтегральна оцінка, яка враховує сумісно усі три виміри сталого розвитку, і, тим самим, відображає взаємозв'язок між трьома нероздільними сферами розвитку суспільства: економічною, екологічною та соціальною. Її структуру для аналізу сталого розвитку країн світу представлено на рис. 4.17 [13], регіонів України – рис. 4.18 [12]. Ступінь гармонізації сталого розвитку відображає баланс між його економічним, екологічним та соціально-інституціональним вимірами.

Оцінки вимірів сталого розвитку, які використані для визначення компоненти якості життя повинні ґрунтуватися на даних про широкий спектр явищ різної природи. Також такі оцінки повинні бути інтегральними, тобто характеризувати певну сторону життєдіяльності людини, як цілісну систему.

З метою кількісного оцінювання вимірів сталого розвитку використовуються принципи побудови ієрархічної системи показників та індексів, які визначаються як L_1 -норми:

$$I_i = \sum_{j=1}^n w_j x_{i,j}, i = \overline{1, m}, \sum_{j=1}^n w_j = 1 \quad (4.4)$$

в просторі показників $X^1 \times X^2 \times \dots \times X^m$, що характеризують економічний, екологічний та соціально-інституціональний розвиток кожної i -ої території (країни, регіону). Вагові коефіцієнти w_j в формулі (4.3) визначаються експертним шляхом.

Використання формули (4.4) потребує узгодження різноманітних даних, як за одиницями вимірювання, так і за діапазоном значень [3, 6].

Тому, якщо більші значення показника X^i відповідають кращому стану сталого розвитку, то застосовується логістичне нормування значень показників за формулою:

$$C_{norm}(x_{i,j}) = \left(1 + e^{\frac{a-x_{i,j}}{b}} \right)^{-1}, \quad (4.4)$$

в якій параметри a і b обчислюються як середнє значення та стандартне відхилення за вибіркою територій, що аналізуються.

В протилежному випадку, коли найбільші значення показника X^i відповідають гіршому стану сталого розвитку, використовується обернена до обрахованої за формулою (4.4) величина:

$$C_{norm}(x_{i,j}) = 1 - \left(1 + e^{\frac{a-x_{i,j}}{b}} \right)^{-1}. \quad (4.5)$$

Сумарний вплив сукупності глобальних загроз на різні території й групи територій (країн, регіонів тощо) оцінюються за допомогою компоненти безпеки життя людей C_{sl} як складової індексу сталого розвитку в формулі (4.1) [12, 13]. Кожній території j ставиться у відповідність вектор:

$$Tr_j = (t_i^j), \quad j = \overline{1, n} \quad (4.6)$$

з координатами $t_i^j \in [0, 1]$, $i = \overline{1, n}$, які характеризують ступінь прояву кожної з загроз.

В якості глобальних загроз виділимо ті, що визначені головними в XXI сторіччі такими авторитетними міжнародними організаціями, як ООН, Всесвітня організація охорони здоров'я (ВООЗ), міжнародні організації «World Economic Forum», «Transparency International», «Global Footprint Network», «International Energy Agency», «World Resources Institute», компанія British Petroleum та іншими: Перелік глобальних загроз і джерел даних, використані в цьому дослідженні, наведено в таблиці 4.6 [13, 153].

Значення показників, що характеризують прояв зазначених загроз, нормуються за допомогою формул (4.4) – (4.5) таким чином, щоб більшому прояву загрози відповідали значення, наближені до 1.

Значення компоненти безпеки життя C_{sl} для j -ої території визначається як норма Мінковського вектора $\bar{S}_j = (s_i^j)$, $s_i^j = 1 - t_i^j, i = \overline{1, n}$:

$$C_{sl} = \|\bar{S}_j\| = \left(\sum_{i=1}^n (s_i^j)^p \right)^{\frac{1}{p}} \quad (4.7)$$

з параметром $p = 3$.

Оскільки збільшення параметра p збільшує відгук (чутливість) моделі на зміни кожної зі складових вектора $\bar{T}r_j$, і навпаки, його зниження згладжує (огрубляє) цю чутливість, в моделі прийнято $p = 3$.

Вводиться також значення вразливості території до сукупності загроз, яке є оберненою величиною до компоненти безпеки життя C_{sl} :

$$I_{vul} = \sqrt[3]{n} - C_{sl}. \quad (4.8)$$

Таким чином, компонента безпеки життя – інтегральна оцінка, яка враховує сумарний вплив сукупності глобальних загроз на сталий розвиток територій, а індекс вразливості території до впливу сукупності загроз відображає ступінь наближення цієї території одночасно до усіх загроз у просторі, який визначається нормою Мінковського [12, 13].

Розглянемо сформульоване завдання на прикладі розрахунку впливу сукупності загроз на територіально-адміністративні одиниці (області та території) України.

У таблиці 4.7 представлені загрози, які були визначені експертним шляхом з урахуванням специфіки сталого розвитку регіонів України [12, 154]. Показники загроз з таблиці 4.7. відрізняються змістом, способом визначення та одиницями вимірювань, тобто мають різну природу. Тому побудова агрегованих (інтегральних) оцінок безпеки була виконана із застосуванням

запропонованого в роботі методу системного узгодження значень цих показників.

За значеннями компонента безпеки життя людей було виконано кластеризацію, що дозволила розподілити регіони України на чотири кластери – з високим, вище середнього, нижче середнього і низьким рівнем безпеки.

Таблиця 4.6.

Глобальні загрози

№	Загроза		Організація	Посилання
	Позначення	Найменування		
1	BB	Biodiversity Balance	Global Footprint Network	http://www.footprintnetwork.org
2	CI	Conflicts Intensity	Heidelberg Institute for International Conflict Research	http://www.hiik.de
3	IG	Information Gap	International Telecommunication Union	http://www.itu.int/
4	CP	Corruption Perception	Transparency International	http://transparency.org
5	ES	Global Reduce of Energy Security	World Bank	http://data.worldbank.org
6	GD	Spread of Global Infectious Diseases	World Bank	http://data.worldbank.org
7	GW	Global Warming	Carbon Dioxide Information Analysis Center	http://data.worldbank.org
8	ND	Vulnerability to Natural Disasters	The International Disaster Database	http://www.emdat.be
9	PI	Proliferation	World Data Center for Geoinformatics and Sustainable Development	http://wdc.org.ua/data
10	WA	Limited Access to Potable Water	WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation	http://data.worldbank.org
11	GINI	Increasing Inequality between People and Nations on Earth	World Bank	http://data.worldbank.org
12	SF	State Fragility	The Fund for Peace	http://ffp.statesindex.org

Таблиця 4.7.

Індикатори загроз для аналізу сталого розвитку регіонів України

№	Індекс	Загроза	Показник	Одиниця вимірювання
1	Соціальна безпека	Зниження тривалості життя	Середня очікувана тривалість життя при народженні по регіонах	роки
2		Злочинність	Коефіцієнт злочинності по регіонах	кількість зареєстрованих злочинів на 100 тис. наявного населення
3		Погіршення здоров'я	Індекс смертності населення від туберкульозу, від новоутворень, від хвороб систем кровообігу, від СНІДу	ум.од
4		Корупція	Індекс сприйняття корупції	%
5		Соціальна нерівність	Індекс Джині (коефіцієнт нерівності в отриманні доходів)	ум.од
6	Економічна безпека	Рівень безробіття	Рівень зареєстрованого безробіття	% на кінець року
7		Технологічна відсталість	Ступінь зносу основних засобів	%
8		Зниження добробуту населення	Наявний дохід населення у розрахунку на одну особу	грн.
9	Екологічна безпека	Забруднення довкілля	Індекс забруднення атмосферного повітря та поверхневих вод	умов. од
10		Дефіцит водних ресурсів	Сукупний обсяг місцевого річкового стоку і запасів підземних вод	тис. м ³ /км ²
11		Вплив на змінювання клімату	Щільність викидів двоокису карбону	т СО ₂ екв./км ²

З використанням методу експоненційного нормування значення всіх показників загроз приведені до безрозмірних величин, які змінюються в діапазоні [0-1]. При цьому значення 0,5 відповідає середньому по країні впливу загрози, а значення, близьке до 1,0 – найбільшому впливу. Визначивши критичний поріг (в даному випадку 0,7) можна виділити критичні значення показників загроз для кожного регіону. На основі застосування співвідношень

4.1 і 4.2 був розрахований компонент безпеки життя людей C_{sl} , наведений в таблиці А.1 додатку А, який агрегує всі 12 загроз, представлених в таблиці 4.6.

У перший кластер з високим рівнем безпеки життя людей ($C_{sl} > 1.1$) увійшло сім регіонів: Івано-Франківська, Чернівецька, Львівська, Тернопільська, Київська, Чернігівська, Хмельницька області. Ці області характеризуються помірним впливом загроз економічного і соціального характеру. При цьому для Київської та Чернігівської областей суттєвою є загроза «Зниження тривалості життя». Крім того, для Київської області важливою загрозою є «Корупція», а для Чернігівської – «Рівень безробіття». Загроза «Зниження добробуту населення» суттєвою є для Чернівецької та Тернопільської областей.

До другого кластеру з рівнем безпеки життя людей вище середнього ($1.1 > C_{sl} > 0.99$) віднесено вісім областей України: Харківська, Черкаська, Волинська, Вінницька, Херсонська, Полтавська, Сумська області та місто Севастополь. При цьому високий рівень загроз: «Злочинність», «Корупція» і «Забруднення навколишнього середовища», характерні для м. Севастополя, для Волинської області істотними загрозами є «Корупція» і «Зниження добробуту населення», для Вінницької і Полтавської областей – «Рівень безробіття» та «Зношеність технологічної інфраструктури», для Сумської області – «Корупція» і «Рівень безробіття».

У третій кластер з рівнем безпеки життя людей ($0.99 > C_{sl} > 0.87$) нижче середнього увійшло шість регіонів України: Закарпатська, Рівненська, Запорізька, Житомирська, Миколаївська області та АР Крим. Для Закарпатської, і Житомирської областей цієї групи одночасно істотними є три загрози, для Миколаївській – чотири.

Четвертий кластер з рівнем безпеки життя людей ($0.87 > C_{sl}$) включає шість регіонів України: Луганську, Одеську, Дніпропетровську, Кіровоградську, Донецьку області та місто Київ. На зниження безпеки життя

в регіонах цієї групи одночасно впливають від трьох (для Луганської області) до семи загроз (для Донецької області).

У додатку А наведено екранні форми, таблиці, графіки і діаграми, які показують як процес проведення дослідження, так і його результати.

Реалізований за допомогою засобів платформи підтримки міждисциплінарних досліджень інструментарій дозволяє досліджувати динаміку індексу сталого розвитку та його складових, здійснювати порівняльний аналіз досліджуваних об'єктів (адміністративно-територіальних одиниць) за широким спектром показників сталого розвитку (рис. 4.17, 4.18, табл. 4.6, 4.7).

В процесі накопичення даних в часовому вимірі та щорічного корегування сценаріїв соціально-економічного розвитку України, побудована он-лайн платформа буде постійно вдосконалюватися. Це дозволить підвищувати достовірність скорегованих сценаріїв та уточнювати стратегії розвитку і рекомендації, які надаються особам, що приймають рішення (ОПР) в режимі сценарного моделювання на усіх рівнях управління державою.

При цьому інтерактивні тематичні панелі, розроблені за допомогою інструментів платформи з генерування проблемно-орієнтованих застосунків на основі інтеграції мікросервісів, можуть бути налаштованими відповідно до вимог ОПР і експертів та відображати результати роботи системи моніторингу виконання сценаріїв, а сам процес сценарного моделювання може здійснюватися в спеціально створених ситуаційних центрах.

4.3.2. Приклад оцінювання ефективності суспільних перетворень з використанням розробленої платформи

Наведемо приклад проведення ще одного актуального дослідження з використанням розробленої платформи – аналізу готовності і оцінювання ставлення населення України до проведення масової вакцинації проти COVID-19.

До оцінювання ефективності масових суспільних заходів (перетворень) застосуємо комбінований підхід, запропонований в роботах [8, 17], який інтегрує методи експертних оцінок, сентимент-аналізу та інтелектуального аналізу текстових повідомлень з відкритих онлайн-джерел і соціальних мереж для формалізованого узгодження результатів.

Значна кількість інформаційних ресурсів в глобальних мережах містить думки людей та їх експертні оцінки. Інформацію, що екстрагується із документів з відкритих джерел, можна аналізувати, узагальнювати, створювати базу для подальшого застосування. Ці дані відрізняються від традиційних експертних оцінок як за обсягами, так і за рівнем об'єктивності. Статистично в соціальних мережах акумулюється 30%-40% інформації, пов'язаної з реакцією користувачів мережі на діяльність органів державної влади.

Виходячи з цього, аналіз інформації з веб-ресурсів повинен мати суттєве значення в державному управлінні як один з каналів зворотного зв'язку.

Для застосування математичних методів прогнозування і передбачення наслідків проведення суспільних перетворень формуються часові ряди, що відповідають обсягам публікацій в соціальних медіа за певні проміжки часу. При цьому найважливіше значення при виконанні контент-аналізу має тональність публікацій.

Головна ідея розроблюваного інструментарію інтелектуального аналізу великих обсягів даних із соціальних медіа – це одночасне застосування методів і засобів інформаційного пошуку, аналізу і агрегування даних із сучасних інформаційних потоків.

Створені в рамках платформи програмні засоби забезпечують автоматичне сканування і первинну обробку інформації з веб-сайтів, соціальних мереж, месенджерів. Використовуються повідомлення із наступних соціальних медіа: Twitter, Youtube, Rutube, Telegram, Facebook, Reddit, Medium, Instagram, LiveJournal, тощо, низки веб-сайтів, які формуються спеціальними програмними модулями, що входять до складу

системи. На основі цього створюються повнотекстові бази даних, виявляються схожі за змістом інформаційні повідомлення, здійснюється повнотекстовий пошук. Відбувається екстрагування понять і проводиться лінгвістичний контент-аналіз текстових повідомлень, визначаються тональності повідомлень на базі застосування нейронних мереж і машинного навчання. Здійснюється аналіз та візуалізація даних, виконуються дослідження динаміки тематичних інформаційних потоків; проводиться прогнозування та передбачення на основі аналізу динаміки публікацій в соціальних медіа.

Суспільні заходи (перетворення), що оцінюються показником суспільних перетворень R , будуть настільки ефективними, наскільки вектор очікувань та соціальної активності суспільства \vec{S} буде за напрямом наближатися до вектора дій влади \vec{G} та наскільки високою і позитивною буде синергія суспільства (рис. 4.19). Синергія суспільства – додаткова суспільна активність (суспільна енергія), яка утворюється в результаті взаємодії громадян країни та може бути позитивною, коли суспільство усвідомлює, що успіхи породжують нові успіхи і від’ємною, коли невдачі призводять до втрати надії людей на позитивні зміни (апатії) і, як наслідок, породжують нові невдачі.

Для визначення показника суспільних перетворень R використовується відома модель:

$$R = |\vec{G}| + |\vec{S}| \cos \alpha + K_s \left(|\vec{G}| + |\vec{S}| \right)^2 \cos \alpha; \alpha \in [0^\circ, 180^\circ] \quad (4.9)$$

Де α – це кут між векторами \vec{S} і \vec{G} , а K_s – коефіцієнт синергії суспільства. Для різних станів суспільства, які можуть змінюватися від суспільного підйому та оптимізму до втоми і апатії, коефіцієнт синергії суспільства K_s може набувати різних значень, зокрема $K_s = [0.25; 0.5; 0.75; 1.0]$.

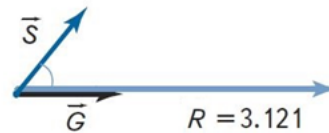
Залежність синергії суспільства від кута α є нелінійною. Граничні значення кута α такі: $\alpha = 0^\circ$, коли напрями вказаних двох векторів збігаються (максимальні суспільні перетворення і максимальна позитивна синергія суспільства $R = 4$, рис. 4.19а), та $\alpha = 180^\circ$, коли їхні напрями

протилежні (відсутність будь яких суспільних перетворень, утворюється максимальна негативна суспільна енергія, яка значно перевищує енергію дій влади, в результаті чого виникає протестний (революційний) стан суспільства $R = -2$, рис. 4.19e). У випадку $\cos \alpha = -1/3$ ($\alpha \approx 109,5^\circ$) влада і суспільство будуть діяти незалежно один від одного (показник суспільних перетворень $R = 0$, рис. 4.19c).

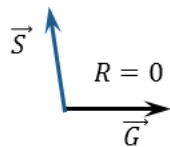


a) $\alpha = 0$; $R = 4$;

максимальні суспільні перетворення

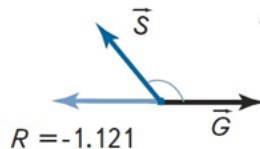


b) $\alpha = 45$; $R = 3.121$;

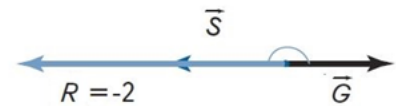


c) $\alpha \approx 109,5^\circ$; $R = 0$;

суспільство і влада діють
незалежно один від
одного, суспільна апатія



d) $\alpha = 135$; $R = -1.121$;



e) $\alpha = 180$; $R = -2$;

відсутність суспільних
перетворень,
виникає революційний
стан суспільства

Рис. 4.19. Ефективність суспільних заходів (масштаб перетворень) при $K_s=0,5$

Для оцінки поляризації думок щодо дій влади стосовно конкретної проблеми вводиться коефіцієнт конфліктності C , який набуває значень у діапазоні $[0,1]$. Граничні значення коефіцієнта конфліктності C такі: $C = 1$ відповідає максимальному рівню поляризації думок у суспільстві; $C = 0$ відповідає максимальному рівню консолідації суспільних настроїв і думок.

У цій моделі (4.9) оцінювання кута α (з подальшим розрахунком показника суспільних перетворень R) може здійснюватися як з використанням методу експертних оцінок (методу Дельфі), так і шляхом автоматичного онлайн-моніторингу суспільних настроїв в інформаційному просторі, або обома способами одразу.

У випадку, коли використовується k інформаційних джерел та потрібно отримати узагальнене значення показника суспільних перетворень, скористаємося наступною формулою:

$$\tilde{R} = \sum_{i=1}^k w'_i R_i; w'_i = \frac{cons_i}{\sum_{j=1}^k cons_j}, \quad (4.10)$$

де \tilde{R} – узагальнене значення показника суспільних перетворень, обчислене на основі даних, отриманих з k інформаційних джерел, R_i – значення показника суспільних перетворень, обчислене на основі даних з i -го інформаційного джерела, $w'_i, \sum_{i=1}^k w'_i = 1$ – ваговий коефіцієнт для значення R_i , пропорційний його узгодженості $cons_i = 1 - C_i$.

У цьому випадку узагальнений коефіцієнт конфліктності \tilde{C} можна визначити як усереднене значення:

$$\tilde{C} = \frac{\sum_{i=1}^k C_i}{k}.$$

Для зручності інтерпретації показника суспільних перетворень скористаємося його нормованим значенням:

$$R_n = \frac{R - R_{\min}}{R_{\max} - R_{\min}},$$

де R_n – нормоване значення показника суспільних перетворень, R – значення показника суспільних перетворень, обчислене за формулою (4.9) або (4.10), R_{\min} та R_{\max} – відповідно, найменше та найбільше значення показника суспільних перетворень.

У додатку Б описано процес організації і проведення дослідження з оцінювання ставлення суспільства до щеплення проти COVID-19 за допомогою інструментарію «PRO ET CONTRA» інтегрованої онлайн-платформи «Advanced Analytics» Світового центру даних з геінформатики та сталого розвитку», який працює на базі КПІ ім. Ігоря Сікорського. Платформу «Advanced Analytics» СЦД-Україна створено на основі розробленої в рамках дисертаційної роботи платформи підтримки міждисциплінарних досліджень.

Продemonстровано приклади роботи інструментів та сервісів розробленої платформи.

Висновки за розділом

1. Розроблена сервіс-орієнтована архітектура платформи підтримки міждисциплінарних досліджень, описані принципи функціонування взаємодії її компонентів.

2. З метою ефективної реалізації розробленої архітектури платформи виконана класифікація її компонентів на дані і власники даних і сервісів, ресурси, які поділяються спільнотою ССД (сервіси даних, сервіси опрацювання даних та управління потоками даних), компоненти СЦД (контейнери ресурсів, менеджер ресурсів, компоненти пошуку і формування сценаріїв), ролі учасників.

3. Розглянуто технічні та технологічні аспекти підтримки міждисциплінарних досліджень, інформаційне середовище СЦД-Україна, технологічний процес інтелектуального опрацювання даних, інтерфейс користувача.

4. Розглянуто методологію, інструментарій, технічні засоби для вирішення складної міждисциплінарної задачі кількісного оцінювання сталого розвитку в глобальному та регіональному контекстах. Представлено результати моделювання процесів сталого розвитку країн світу та адміністративних регіонів України в контексті якості та безпеки життя людей, які можуть використовуватись при прийнятті управлінських рішень на державному, регіональному та муніципальному рівні.

5. Розроблені в рамках платформи програмні інструменти для інтелектуальної обробки, аналізу та системного узгодження даних різної природи, їх систематизації, оцінювання, аналізу якості, коректності, продемонстровані на прикладі вирішення задачі з оцінювання ефективності масових суспільних заходів (перетворень), зокрема оцінювання ставлення населення України до проведення масової вакцинації проти COVID-19.

ВИСНОВКИ

У дисертаційній роботі розв'язано актуальну науково-практичну задачу організації, планування і здійснення підтримки міждисциплінарних досліджень в ССД. Проведені дослідження дозволили отримати наступні наукові результати, які мають істотні переваги перед існуючими рішеннями:

1. Виконано аналіз діяльності Світової системи даних і окремого Світового центру даних як об'єкта управління, виявлено і проаналізовано невирішені проблеми підтримки міждисциплінарних досліджень, обґрунтовано перспективність створення платформи підтримки міждисциплінарних досліджень, як універсального інструментарію вирішення проблем організації діяльності Світової системи даних.

2. *Вперше створено* платформу підтримки міждисциплінарних досліджень в Світовій системі даних, яка забезпечує інтеграцію застосунків і джерел даних, оброблення запитів до розподілених джерел, системне узгодження даних різної природи, проблемно-орієнтоване інтелектуальне оброблення даних та автоматизоване генерування застосунків.

3. *Вперше створено* метод узгодження даних, який відрізняється застосуванням оцінок інформаційних втрат процедур перетворення даних та оцінок узгодженості даних, отриманих з різних джерел, для конструювання процедури перетворення даних з найменшими інформаційними втратами.

4. *Модифіковано* клаузальну логіку планування взаємодії застосунків для розв'язання проблем міждисциплінарних досліджень, яка відрізняється від відомих використанням системи типів, орієнтованих на організацію взаємодії застосунків, комплексом спеціальних аксіом описання застосунків на основі передумов та постулов і правил виведення, які визначають принципи комплексування застосунків, що дозволяє у зручному для користувача вигляді визначати проблему і будувати схему її вирішення.

5. *Отримав подальший розвиток* метод виведення в клаузальній логіці взаємодії застосунків за рахунок попереднього виведення в просторі типів з

наступною деталізацією виводу в системі індивідних об'єктів та відновленням схеми розв'язання проблеми на основі виводу.

6. Компоненти розробленої платформи підтримки міждисциплінарних досліджень в Світовій системі даних знайшли впровадження в Інституті проблем реєстрації інформації НАН України, Інституті геофізики НАН України ім. С.І. Субботіна, Інформаційно-аналітичному ситуаційному центрі КПІ ім. Ігоря Сікорського, Навчально-науковому комплексі «Світовий центр даних з геоінформатики та сталого розвитку» КПІ ім. Ігоря Сікорського, Корпорації «Науковий парк «Київська Політехніка» та інших організаціях, що дозволило підтвердити їх працездатність, скоротити проектування і реалізацію застосувань на 20 – 30 % в залежності від особливостей діяльності, скоротити втрати внаслідок невчасного отримання потрібних даних на 11,75 %, скоротити інформаційні втрати при перетворенні даних на 7,25%, підвищити обсяги наданих користувачам інформаційних ресурсів на 50%, підвищити ефективність аналізу даних різної природи експертами без знання інструментальних мов програмування на 42%.

7. Розроблені методичні та програмні засоби використовуються на кафедрі математичних методів системного аналізу та кафедрі автоматики та управління в технічних системах КПІ ім. Ігоря Сікорського при підготовці фахівців в галузі інтелектуального аналізу даних та інформаційних технологій.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. E. Ziemba, "The ICT Adoption in Government Units in the Context of the Sustainable Information Society", 2018 Federated Conference on Computer Science and Information Systems, pp.725–733. DOI: 10.15439/2018F116.
2. M. Z. Zgurovsky, A. D. Gvishiani, K. V. Yefremov and A. M. Pasichny. Integration of the Ukrainian science into the world data system // *Cybernetics and Systems Analysis: Volume 46, Issue 2 (2010)*, pp. 211-219.
3. Теленик С.Ф., Амонс О.А., Єфремов К.В., Лиско В.Т. Логічний підхід до інтеграції програмних застосувань підтримки міждисциплінарних наукових досліджень // *Наукові вісті НТУУ «КПІ»*. – 2013. – № 5. – С. 53-72.
4. Telenyk, S., Nowakowski, G., Yefremov, K., Khmeliuk, V. Logics based application integration for interdisciplinary scientific investigations. *Proceedings of the 2017 IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2017*, P. 1026-1031.
5. Теленик С.Ф., Амонс О.А., Єфремов К.В., Жук С.В. Семантична інтеграція різномірних інформаційних ресурсів // *Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: Зб. наук. пр.* – К.: Век+, – 2013. – No. 58. – с.29-45.
6. M.Z. Zgurovsky, A.A. Boldak, K.V. Yefremov. Intelligent analysis and the systemic adjustment of scientific data in interdisciplinary research // *Cybernetics and Systems Analysis: Volume 49, Issue 4 (2013)*, pp. 541-552.
7. Болдак А.О., Єфремов К.В. Предметно-орієнтована мова аналітичної обробки даних // *Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: збірник наукових праць.* – К.: Век+, – 2012. – No. 55. – с.67-71.
8. Foresight and construction of the strategies of socio-economic development of Ukraine on mid-term (up to 2020) and long-term (up to 2030) time horizons / Scientific advisor of the project acad. of NAS of Ukraine M. Zgurovsky //

International Council for Science (ICSU); Committee for the System Analysis of the Presidium of NAS of Ukraine; National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»; Institute for Applied System Analysis of MES of Ukraine and NAS of Ukraine; World Data Center for Geoinformatics and Sustainable Development; Agrarian Superstate Foundation. – 2nd ed. – Kyiv : NTUU «Igor Sikorsky KPI», Publ. house «Polytechnica», 2016. – 184 p.

9. Єфремов К. В., Мазур Р.Ф. Системний та структурний рівні організації мультиагентних систем з можливістю адаптації процесу обробки даних // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: Зб. наук. пр. – К.: Век+, – 2011. – № 53. – с. 6–9.
10. Nowakowski, G., Telenyk, S., Yefremov, K., Khmeliuk, V. The approach to applications integration for world data center interdisciplinary scientific investigations. Proceedings of the 2019 Federated Conference on Computer Science and Information Systems. FedCSIS 2019, P. 539-545.
11. Marsel Shaimardanov, Alexei Gvishiani, Michael Zgurovsky, Alexander Sterin, Alexander Kuznetsov, Natalia Sergeyeva, Evgeny Kharin, Kostiantyn Yefremov. Development of WDS Russian-Ukrainian segment // Data Science Journal. – 2013. – Volume 12. – p. 17-26.
12. Sustainable Development Analysis: Global and Regional Contexts / International Council for Science (ICSU) and others; Scientific Supervisor of the Project M. Zgurovsky. — K. : Igor Sikorsky KPI, 2017. — Part 2. Ukraine in Sustainable Development Indicators (2016–2017). — 72 p.
13. Sustainable Development Analysis: Global and Regional Contexts / International Council for Science (ICSU) and others; Scientific Supervisor of the Project M. Zgurovsky. — K. : Igor Sikorsky KPI, 2017. — Part 1. Global Analysis of Quality and Security of Life (2016). — 208 p.
14. Zgurovsky M., Boldak A., Melnyk O., Perestyuk M., Putrenko V., Pyshnograiev I., Yasinsky V., Yefremov K., Foresight 2018: systemic world conflicts and global forecast for XXI century / International Council for

- Science etc.; Scientific Supervisor M. Zgurovsky. – K. : NTUU «Igor Sikorsky Kyiv Polytechnic Institute», 2018. — 226 p.
15. Згуровський М.З., Болдак А.О., Джигирей І.М., Єфремов К.В., Форсайт 2018: Аналіз підготовки і перепідготовки фахівців природничого і технічного спрямування, виходячи з цілей сталого соціально-економічного розвитку України до 2025 року. — К. : НТУУ «КПІ імені Ігоря Сікорського», Вид-во «Політехніка», 2018. – 32 с.
 16. Згуровский М.З., Болдак А.А., Ефремов К.В., Сергеева Н.А., Забаринская Л.П., Шестопапов И.П., Нисилевич М.В. Применение методов интеллектуального анализа данных для эмпирических исследований взаимосвязи гелио- и геофизических процессов // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: збірник наукових праць. – К.: Век+, – 2013. – No. 58. – с.4-10.
 17. M. Zgurovsky, A. Boldak, D. Lande, K. Yefremov and M. Perestyuk. Predictive Online Analysis of Social Transformations based on the Assessment of Dissimilarities between Government Actions and Society's Expectations. Conference proceedings of 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2020, P. 130-135..
 18. Zgurovsky, M., Putrenko, V., Dzhygyrey, I., Boldak, A., Yefremov, K., Pashynska, N., Pyshnograiev, I., Nazarenko, S. Parameterization of Sustainable Development Components Using Nightlight Indicators in Ukraine. Conference proceedings of 2018 IEEE 1st International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2018, P. 1-5.
 19. Zgurovsky, M., Boldak, A., Yefremov, K., Pyshnograiev, I. Modeling and investigating the behavior of complex socio-economic systems. 2017. IEEE 1st Ukraine Conference on Electrical and Computer Engineering, UKRCON 2017 – Proceedings, P. 1113-1116.
 20. Павлов А.О., Теленик С.Ф. Алгоритмизация и ИТ в управлении. Киев: Техника, 2002 – 320 с.
 21. Силкин Б. И., Троицкая В. А., Шебалин Н. В. Наша незнакомая планета.

- М.: Изд-во АН СССР, 1962. – 295 с.
22. Minster J.-B., Capitaine N., Clark D. M. The new ICSU World Data System: A world-wide community of excellence for data issues // Materials of the International Conference "Electronic Geophysical Year: State of the Art and Results", Ed. by V. Nechitailenko, GC RAS. – Moscow, 2009. – 117p., doi:10.2205/2009-REGYconf.
 23. Zgurovsky M., Yefremov K., Pasichny A. WDC-Ukraine as a part of Russian-Ukrainian WDC Segment // Materials of the International Conference "Electronic Geophysical Year: State of the Art and Results", Ed. by V. Nechitailenko, GC RAS. – Moscow, 2009. – 117p., doi:10.2205/2009-REGYconf.
 24. М. Згуровский М.З., Патон Б.Е., Якименко Ю.И. Состояние и перспективы развития национальной телекоммуникационной академической сети, 1997, <http://www.uazone.org/inet/uren/uran-dop1w97.html>
 25. Yakymenko Yu., Timofeyev V., Galagan V., Dombrougov M. Development and European Integration of Ukrainian Research and Academic Network (URAN) for Provision of High Speed Services to Science and Education // Materials of the 21st Int. CODATA Conf., 2008, Kiev, Ukraine
 26. Guide to the World Data Center System: General Principles – World Data Centers – Data Services. – Paris.: ICSU, 1996. – 110 p.
 27. A.Gvishiani, M.Zgurovsky, V.Starostenko, K.Yefremov, A.Pasichny, N.Sergeyeva. World Data Center for Geoinformatics and Sustainable Development: state of the art. // Proceedings of the 21st Int. CODATA Conf., 2008, Kiev, Ukraine
 28. Kateryna Pereverza, Kostiantyn Yefremov, Alexei Pasichny, Alexei Leonov. Methodology of scenario development based on Ukrainian branch of the WDC interdisciplinary data. // Proceedings of the 21st Int. CODATA Conf., 2008, Kiev, Ukraine
 29. Starostenko V., Yatskiv Ya., Lyalko V., Ivanov V., Rudenko L., Yefremov K.

Ukrainian science data: mutual goals and approaches // Materials of the 21st Int. CODATA Conf., 2008, Kiev, Ukraine

30. Системи управління хмарними ІТ-структурами / С.Ф. Теленик, О.І. Ролік, М.В. Ясочка, О.С. Квітко // Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту: Матеріали Міжнар. наук. конф., 16 – 20 травня 2011 р., м. Євпаторія. – Херсон: ХНТУ, 2011. – 1. – С. 124-127.
31. Вейдж Д., Миллер Х.Г. Вычисления в облаке: будут ли полезны открытые сервисы? // Открытые системы. – 2010. – №1. – С.68-75.
32. S. El-Seoud, H. El-Sofany, M. Abdelfattah, M. Reham, “Big data and cloud computing: trends and challenges,” International Journal of Interactive Mobile Technologies, vol. 11, issue 2, pp. 34-52, 2017.
33. M. Kavis, Architecturing the Cloud, Wiley, 2014, 199 p.
34. Когаловский М. Р. Интеграция данных в информационных системах //Сб. трудов Третьей Всероссийской конференции “Стандарты в проектах современных информационных систем”, Москва. – 2003. – С. 23-24.
35. Development of the WMO Core Profile of the ISO Metadata standard [Електронний ресурс] – Режим доступу: <http://www.wmo.int/pages/prog/www/WDM/Metadata/documents.html>. – Назва з екрану.
36. SeaDataNet Metadata formats [Електронний ресурс] – Режим доступу: <http://www.seadatanet.org/Standards-Software/Metadata-formats>. – Назва з екрану.
37. Levy A.Y. Logic-Based Techniques in Data Integration. Logic-based Techniques in Data Integration. In: Logic Based Artificial Intelligence. Edited by J. Minker. Kluwer Publishers, 2000.
38. Kossmann D. The state of the art in distributed query processing. – New York: ACM Computing Surveys vol.32, «ACM», 2000. – P. 422-469.
39. ANSI/X3/SPARC Study Group on Data Base Management Systems Interim Report. FDT Bulletin, 7 (2), 1975, pp. 1-140.
40. Тузовский А.Ф. Интеграция баз данных на основе онтологий [Електронний ресурс]. – Режим доступу:

ontology.ipi.ac.ru/files/d/dc/DBIntegration.ppt. – Назва з екрану.

41. Имхофф К. Аналитические решения: понимание трех составляющих интеграции - EAI, EII, ETL (перевод с английского) [Электронный ресурс]. – Режим доступа: <http://citforum.edunet.kz/consulting/BI/integration/>. – Назва з екрану.
42. Wu T. EII – ETL – EAI - What, Why, and How! [Электронный ресурс]. – Режим доступа: https://www6.software.ibm.com/developerworks/tw/events/20051028/db2_6b.pdf. – Назва з екрану.
43. Wiederhold G. Mediators in the Architecture of Future Information Systems / G.Wiederhold // IEEE Computer. –1992. – Vol.25. – #3. – P. 38-49.
44. J Ullman. Principles of Data and Knowledge Base Systems. Volume I. Computer Science Press, Woodland Hills, CA, 1988.
45. Cattell, R. G., Barry, D. K., Berler, M., Eastman, J., Jordan, D., Russell, C., Schadowa, O., Stanienda, T., Velez, F. (2000). The Object Data Management Standard: ODMG 3.0. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.. ISBN: 1-55860-463-4.
46. Abiteboul S. Querying semi-structured data. In: Afrati F., Kolaitis P. (eds) Database Theory — ICDT '97. ICDT 1997. 1996. Lecture Notes in Computer Science, vol 1186. Springer, Berlin, Heidelberg.
47. Peter Buneman. Semistructured data. In Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems (PODS '97). 1997. Association for Computing Machinery, New York, NY, USA, 117–121. DOI:<https://doi.org/10.1145/263661.263675>
48. Calvanese D., De Giacomo G., Lembo D., Lenzerini M., Rosati R. Data Management in Peer-to-Peer Data Integration Systems [Электронный ресурс]. – Режим доступа: <http://www.dis.uniroma1.it/~degiacono/papers/2006/calv-et-al-GDM-book-06.pdf>. – Назва з екрану.
49. Кузнецов С.Д. Основы современных баз данных [Электронный ресурс]. – Режим доступа: <http://www.citforum.ru/database/osbd/-contents.shtml>. – Назва з екрану.

50. Интеграция данных: синтаксис и семантика / Л.Черняк // Открытые системы. – 2009. – №10. [Электронный ресурс]. – Режим доступа: <http://www.osp.ru/os/2009/10/11170978/>. – Назва з екрану.
51. Розробка онтології матеріалознавства засобами Protégé-OWL / Д.Г. Досин, Р.Р. Даревич, Н.В. Шкутяк // Штучний інтелект. – 2008. – № 3. – с. 70-77.
52. Д.И. Муромцев. Онтологический инжиниринг знаний в системе Protégé. – СПб: СПб ГУ ИТМО, 2007. – 62 с.
53. Шаповалова С.И., Ефремов К.В., Глуханик А.И. Организация интегрированного доступа к информационным ресурсам [Текст] / С.И. Шаповалова, К.В. Ефремов, А.И. Глуханик / – Сборник трудов конференции ИАИ-2011 – 2011. – С.102-108.
54. Пол Спенсер. XML. Проектирование и реализация. М.:Лори.–2001.–510 с.
55. OWL Web Ontology Language W3C Recommendation [Электронный ресурс]. – Режим доступа: <http://www.w3.org/TR/owl-features/>. – Назва з екрану.
56. I. Horrocks et al. From SHIQ and RDF to OWL: the making of a Web Ontology Language / Web Semantics: Science, Services and Agents on the World Wide Web 1 (2003), p. 7–26.
57. DF Primer W3C Recommendation [Электронный ресурс]. – Режим доступа: <http://www.w3.org/TR/rdf-primer/>. – Назва з екрану.
58. Жигалов В. А. Технология построения естественно-языковых интерфейсов к структурированным источникам данных – М.: Диссертация на соискание ученой степени канд. техн. наук, 2000. – 172 с.
59. SPARQL Query Language for RDF [Электронный ресурс]. – Режим доступа: <http://www.w3.org/TR/rdf-sparql-query/>. – Назва з екрану.
60. ARQ - A SPARQL Processor for Jena [Электронный ресурс]. – Режим доступа: <http://jena.sourceforge.net/ARQ/>. – Назва з екрану.
61. Quilitz B., Leser U. Querying Distributed RDF Data Sources with SPARQL – Berlin: Lecture Notes in Computer Science vol. 5021, «Springer Berlin Heidelberg», 2008., – с. 524-538.

62. Langegger A., Wöß W., Blöchl M. SemWIQ – Semantic Web Integrator and Query Engine, – München: Informatik 2008 Beherrschbare Systeme dank Informatik, 2008. – с. 718-722.
63. A. Schwarte, P. Haase, K. Hose, R. Schenkel, M. Schmidt. FedX: A Federation Layer for Distributed Query Processing on Linked Open Data. In ESWC Poster and Demo Session Proceedings. Springer, 2011.
64. Kikuchi S., Sachdeva S., Bhalla S. et al., Adaptive Integration of Distributed Semantic Web Data // Berlin: Databases in Networked Information Systems, Volume 5999, «Springer Berlin Heidelberg», 2010. P. 174-193.
65. PANGAEA. Data Publisher for Earth & Environmental Science [Электронный ресурс]. – Режим доступа: <http://www.pangaea.de/>. – Назва з екрану.
66. Diepenbroek, M. , Grobe, H. and Schindler, U. (2008): PANGAEA® - platform for an ICSU World Data Center as a networked publication and library system for geoscientific data , DFG/DINI workshop, Berlin.
67. Franklin M. From Database to Dataspace: A New Abstraction for Information Management / M.Franklin, A.Halevy, D.Maier // ACM SIGMOD. – 2005. – December (Record 34, no. 4). [Электронный ресурс]. – Режим доступа: <http://portal.acm.org/wdc/citation.cfm?id=1107502>. – Назва з екрану.
68. Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. IEEE Intelligent Systems, March–April 2009, pp. 8–12.
69. Halevy A. Information Integration / Alon Halevy // Liu L., Özsu M.T. (eds) Encyclopedia of Database Systems. / Alon Halevy. – New York: Springer-Verlag New York, 2018. – (2).
70. Skyrius R. Business Intelligence. A Comprehensive Approach to Information Needs, Technologies and Culture / R. Skyrius., 2021. – (Progress in IS; 1).
71. C. Labadie, C. Legner, M. Eurich and M. Fadler, "FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs," 2020 IEEE 22nd Conference on Business Informatics (CBI), Antwerp, Belgium, 2020, pp. 201-210, doi: 10.1109/CBI49978.2020.00029.
72. T. Revathi, et al. Hadoop History and Architecture // Big Data Processing With

- Hadoop. – IGI Global, 2019, pp.32-44.
73. Papotti P., Santoro D. (2018) Data Integration. In: Sakr S., Zomaya A. (eds) Encyclopedia of Big Data Technologies. Springer, Cham. https://doi.org/10.1007/978-3-319-63962-8_6-1
 74. A best-effort integration framework for imperfect information spaces. Int. J. Intell. Inf. Database Syst. 11, 4 (January 2018), 296–314.
 75. Шаховська Н.Б. Методи опрацювання консолідованих даних за допомогою просторів даних // Проблеми програмування. – 2011. – № 4. – С. 72-84.
 76. Shakhovska N., Medykovsky M., Stakhiv P. Application of algorithms of classification for uncertainty reduction //Przegląd Elektrotechniczny. – 2013. – V. 89. – pp. 284-286.
 77. Veres, Oleh, Shakhovska, Natalya. Elements of the formal model big date // 2015 XI International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH). – IEEE. – 2015. – pp. 81-83.
 78. Bresciani P., Giorgini P., Giunchiglia F. Tropos: An Agent-Oriented Software Development Methodology – Trento : Kluwer Academic Publishers, 2003.
 79. Padgham L. The Prometheus Methodology / Padgham L., Winikoff M. // Engineering Applications of Artificial Intelligence. - 2004. - 18(2).
 80. Carole B. ADELFE, a Methodology for Adaptive Multi-Agent Systems Engineering : ESAW'02 Proceedings of the 3rd international conference on Engineering societies in the agents world III / Carole B., Sylvain P., Gauthier P. - Berlin : Springer-Verlag, 2003.
 81. Wooldbridge M. The Gaia Methodology for Agent-Oriented Analysis and Design / Wooldbridge M., Jennings N. R., Kinny D. // Autonomous Agents and Multi-Agent Systems. - Kluwer Academic Publishers, 2000. - 3. - pp. 285-312.
 82. Caire F. JADE: A software framework for developing multi-agent applications. Lessons learned / Caire F., Poggi A., Rimassa G. // Information and Software Technology. – 2008. – 50. – pp. 10-21.
 83. Erl T. Service-oriented architecture: concepts, technology, and design. -

- Prentice Hall Professional Technical Reference, 2005.
84. UDDI Version 3.0.2 Specification [Электронный ресурс]. – Режим доступа: http://uddi.org/pubs/uddi_v3.htm. – Назва з екрану.
 85. Web Services Description Language (WSDL) 1.1 [Электронный ресурс]. – Режим доступа: <http://www.w3.org/TR/wsdl>. – Назва з екрану.
 86. J. Greer, Web Services Description Language: 55 Most Asked Questions: What You Need to Know, Emereo Publishing, 2014.
 87. SOAP Version 1.2 Part 1: Messaging Framework (Second Edition) [Электронный ресурс]. – Режим доступа: <http://www.w3.org/TR/soap12-part1/>. – Назва з екрану.
 88. S. Graham, et all, Building Web Services with Java: Making Sense of XML, SOAP, WSDL, and UDDI (2nd Edition), Sams Publishing; 2 edition, 2004.
 89. J. Laznik, Y. Mannari, R. Dhruv, BPEL and Java Cookbook: Over 100 Recipes to Help You Enhance Your SOA Composite Applications with Java and BPEL, Birmingham, 2013.
 90. About the common object request broker architecture specification version 3.3. [Электронный ресурс] – Режим доступа: <https://www.omg.org/spec/CORBA>. – Назва з екрану.
 91. G. Nowakowski, “Rest Api safety assurance by means of HMAC mechanism”, Information Systems in Management, Vol. 5, No. 3, pp. 358-369, 2016.
 92. Fielding, Roy Thomas. Architectural Styles and the Design of Network-based Software Architectures. Doctoral dissertation, University of California, Irvine, 2000.
 93. Windows Communication Foundation Architecture Overview: [Электронный ресурс]. – Режим доступа: <http://msdn.microsoft.com/en-us/library/aa480210.aspx>. – Назва з екрану.
 94. A. Youssef, et al., "Rover: Scalable Location-Aware Computing" in Computer, vol. 36, no. 10, pp. 46-53, 2002.
 95. WS Choreography Model Overview [Электронный ресурс] – Режим доступа: <https://www.w3.org/TR/ws-chor-model/>. – Назва з екрану.

96. Thomas Erl, Service-Oriented Architecture: Concepts, Technology, and Design. 2005. Prentice Hall PTR, USA.
97. Matjaz Juric. BPEL and Java [Електрон. Ресурс]. – Режим доступу: <http://www.theserverside.com/news/1364554/BPEL-and-Java>. – Назва з екрану.
98. Workflow Management System Taverna [Електрон. Ресурс]. – Режим доступу: <http://www.taverna.org.uk/>. – Назва з екрану.
99. A. T. Guler, C. Waaijer, et al, “Automating bibliometric analyses using Taverna scientific workflows: A tutorial on integrating Web Services,” in Journal of Informetrics, 2016.
100. Zgurovsky M., Yefremov K. Formation of a common information space of WDC-Ukraine partners' network // Materials of the International Conference "Electronic Geophysical Year: State of the Art and Results", Ed. by V. Nechitailenko, GC RAS. – Moscow, 2009. – 117p., doi:10.2205/2009-REGYconf
101. Теленик С.Ф. Каталогизация и интеграция разнородных информационных ресурсов / С.Ф.Теленик, С.В. Жук, В.Т. Лыско, К.В. Ефремов // Молодой ученый. –№5(52). –2013. –С.176–179.
102. Linked Data: Evolving the Web into a Global Data Space // Synthesis Lectures on the Semantic Web: Theory and Technology. [2011]. [Електрон. Ресурс]. – Режим доступу: <http://linkeddatabook.com/editions/1.0/>. – Назва з екрану.
103. Kostiantyn Yefremov. Agent-oriented approach for integration of WDC-Ukraine partner network resources // Materials of the 22nd Int. CODATA Conf., 2010, Cape Town, South Africa
104. Manolescu I., Florescu D., Kossman D. Answering XML Queries over Heterogeneous Data Sources. Proc. Of the 27th VLDB Conference, Roma, Italy, 2001.
105. Петренко А.І. Grid та інтелектуальна обробка даних Data Mining / А.І.Петренко // Систем. дослідж. та інформ. технології. — 2008. — № 4. — С. 97-110.

106. R. Queiroz, et al, Functional Interpretation of Logical Deduction, Advances in Logic, v. 5. Singapore: World Scientific, 2012.
107. Jena Toolkit For Semantic Web Applications [Электронный ресурс]. – Режим доступа: <http://jena.apache.org/>. – Назва з екрану.
108. Evren Sirin, Bijan Parsia and oth. Pellet: A practical OWL-DL reasoner — Web Semantics: Science, Services and Agents on the World Wide Web, Volume 5, Issue 2, June, 2007, P. 51-53.
109. The D2RQ Plattform: Accessing Relational Databases as Virtual RDF Graphs [Электронный ресурс]. – Режим доступа: <http://d2rq.org/>. – Назва з екрану.
110. Бездушный А.А. Математическая модель системы интеграции данных на основе онтологий // Вестн. НГУ: Сер. Информационные технологии. Новосибирск, 2008. Т. 6. Вып. 2. с. 15–40.
111. Perrin D. An agent-based approach to immune modelling / Perrin D., Ruskin H. J., Burns J., Crane M. // Lecture Notes in Computer Science. – Springer, 2006. – 3980. – pp. 612-621.
112. Wolfe S. Comparing Route Selection Strategies in Collaborative Traffic Flow Management: IEEE/WIC/ACM International Conference on IAT / Wolfe S., Jarvis P., Enomoto F., Sierhuis M. – 2007.
113. Russell S. J. Artificial Intelligence: A Modern Approach / Russell S. J., Norvig P. – Prentice Hall, 2009.– 3.
114. Wooldridge M. Intelligent agents: theory and practice / Wooldridge M., Jennings N. R. // The Knowledge Engineering Review. – 1995. – 10(2). - pp. 115-152.
115. D'Inverno M. Understanding agent systems / D'Inverno M., Luck M., Luck M. M. – Springer, 2004. – 2 : p. 240.
116. Robertson R. Self-adaptive software: applications: Second International Workshop, IWSAS 2001 (Balatonfüred, Hungary, May 17-19, 2001) / Robertson R., Shrobe H. – Springer, 2003. – Vol. 2.
117. FIPA specifications // Foundation for Intelligent Physical Agents. [2005] [Электронный ресурс]. – Режим доступа:

- <http://www.fipa.org/repository/bysubject.html>. – Назва з екрану.
118. OWL Web Ontology Language // W3C Recommendation. [2009]. [Електронний ресурс]. – Режим доступу: <http://www.w3.org/TR/owl2-overview/>. – Назва з екрану.
 119. DBpedia – A Crystallization Point for the Web of Data // Christian Bizer, Jens Lehmann. [2009]. [Електронний ресурс]. – Режим доступу: www.wiwiiss.fu-berlin.de/en/institute/pwo/bizer/research/publications/Bizer-et-al-DBpedia-CrystallizationPoint-JWS-Preprint.pdf. – Назва з екрану.
 120. Faceted Wikipedia Search // Rasmus Hahn, Christian Bizer. [2010]. [Електронний ресурс]. – Режим доступу: www4.wiwiiss.fu-berlin.de/bizer/pub/hahn-et-al-faceted-wikipedia-search-BIS2010.pdf. – Назва з екрану.
 121. Айвазян С.А. и др. Прикладная статистика: Исследование зависимостей: Справ, изд. / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин; Под ред. С. А. Айвазяна. – М.: Финансы и статистика, 1985. – 487 с.
 122. Somerville M., Rapport D. Transdisciplinarity: reCreating Integrated Knowledge. — Oxford (UK): EOLSS Publ. Co. Ltd., 2000 — 271 p.
 123. Аналіз соціально-економічних процесів розвитку суспільства, заснованого на знаннях / наук. кер. проекту М. З. Згуровський. – К.: КПП ім. Ігоря Сікорського, 2017. – 26 с. ISBN 978-966-622-819-5
 124. M. Zgurovsky, K. Yefremov, M. Perestyuk, V. Putrenko and I. Pyshnograiev, "Foresight COVID-19: Impact on Economy and Society – European Context," Conference proceedings of 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2020, pp. 8-13, doi: 10.1109/SAIC51296.2020.9239211.
 125. Айвазян С. А., Бухштабер В. М., Енюков И. С. и др. Прикладная статистика. Классификация и снижение размерности. – М.: Финансы и статистика, 1989. – 607 с.
 126. Luce R.D., Krantz D.H., Suppes P., Tversky A. Foundation of measurement Vol. 3: Representation, axiomatization and invariance. – San Diego: Academic Press, 1990. – 368 p.

127. Е.С. Вентцель. Теория вероятностей. – М : Высшая школа, 1999. – 576 с.
128. P. Velleman, L. Wilkinson, Nominal, ordinal, interval, and ratio typologies are misleading. 1993. Am. Stat. 47, 65–73.
129. Згуровский М.З., Болдак А.А. Системное согласование данных разной природы в мультидисциплинарных исследованиях // Кибернетика и системный анализ. – 2011. – 46, №5. — С. 152-163.
130. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. – 7-е изд. – М.: Физматлит, 2004. – 572 с.
131. Conway J.H., Smith D. On Quaternions and Octonions. – A K Peters/CRC Press - Abington, Oxfordshire, UK, 2003 – 159 p.
132. Ильин В. А., Позняк Э. Г. Линейная алгебра : учебник для вузов / Ильин В. А., Позняк Э. Г. - 6-е изд. – М.: Физматлит, 2007. – 278 с.
133. Шеннон К. Работы по теории информации и кибернетике. – М.: Изд. иностр. лит., 2002. – 836 с.
134. Balakrishnan N. Handbook of the Logistic Distribution. – NY: Marcel Dekker, 1992. – 601 p.
135. Krippendorff K. Content analysis: An introduction to its methodology / Thousand Oaks, CA: Sage . – 2004. – p. 219-250.
136. Create & manage data. [Электронный ресурс]. – Режим доступа: <http://www.data-archive.ac.uk/create-manage/life-cycle/>. – Назва з екрану.
137. Shearer C. The CRISP-DM model: the new blueprint for data mining // J Data Warehousing. – 2000. – №5. – P. 13-22.
138. Rohanizadeh, S. S. and Moghadam, M. B. A Proposed Data Mining Methodology and its Application to Industrial Procedures // Journal of Industrial Engineering. – 2009. – № 4. – P. 37-50.
139. Constitution of the International Council for Science World Data System (ICSU WDS), 2012. [Электронный ресурс] – Режим доступа: http://icsu-wds.org/images/files/WDS_Constitution_04_04_12.pdf. – Назва з екрану.
140. Jean Bernard Minster. The ICSU World Data System as a Global System of Data Systems // Abstract Proceedings of the XXVth IUGG General Assembly

“Earth on the Edge – Science for a Sustainable Planet”, 2011, Melbourne, Australia

141. ICSU Annual Report 2008, 2008. – Режим доступу: <http://www.icsu.org/publications/annual-reports/annual-report-2008/annual-report-2008-file>. – Назва з екрану.
142. Olsen L.M., G. Major, K. Shein, J. Scialdone, R. Vogel, S. Leicester, H. Weir, S. Ritz, T. Stevens, M. Meaux, C.Solomon, R. Bilodeau, M. Holland, T. Northcutt, R. A. Restrepo. GCMD's Science Keywords and Associated Directory Keywords. – 2007. – P.26.
143. Worldwide Global StorageSphere Forecast, 2020–2024: Continuing to Store More in the Core. [Електронний ресурс] – Режим доступу: <https://www.idc.com/getdoc.jsp?containerId=US46224920>. – Назва з екрану.
144. Dwight Davis, Leading the Data Intelligence Transformation, Jul 9, 2019. [Електронний ресурс] – Режим доступу: <https://www.cio.com/article/3406806/ai-unleashes-the-power-of-unstructured-data.html>. – Назва з екрану.
145. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'13). Curran Associates Inc., Red Hook, NY, USA, 3111–3119.
146. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5, 135-146, <https://arxiv.org/abs/1607.04606>
147. Tools for named entity recognition [Електронний ресурс] – Режим доступу: <https://www.clarin.eu/resource-families/tools-named-entity-recognition>. – Назва з екрану.
148. Аналіз сталого розвитку — глобальний і регіональний контексти: У 2 ч. / Міжнар. рада з науки (ICSU) [та ін.]; Виконавці: А. О. Болдак, С. В. Войтко, І. М. Джигирей та інші : наук. кер. М. З. Згуровський. — К. :

- НТУУ «КПІ», 2009. — Ч. 2. Україна в індикаторах сталого розвитку. Аналіз — 2009. — 200 с.
149. Аналіз сталого розвитку — глобальний і регіональний контексти: У 2 ч. / Міжнар. рада з науки (ICSU) [та ін.]; Виконавці: А. О. Болдак, С. В. Войтко, І. М. Джигирей та інші : наук. кер. М. З. Згуровський. — К.: НТУУ «КПІ», 2009. — Ч. 1. Глобальний аналіз якості та безпеки життя людей. — 2009. — 280 с.
 150. Болдак А. О., Лазаренко І. С. Алгебраїчний підхід до верифікації моделей складних систем на прикладі побудови індексу нерозповсюдження зброї // Системний аналіз та інформаційні технології: матеріали 19-ї Міжнародної науково-технічної конференції SAIT 2017, Київ, 22–25 травня 2017 р. / ННК «ІПСА» НТУУ «КПІ ім. Ігоря Сікорського». — К.: ННК «ІПСА» НТУУ «КПІ», 2017. — С. 41–42.
 151. Zgurovsky M. Z., Pankratova N. D. System Analysis: Theory and Applications//Springer-Verlag, Berlin Heidelberg, 2007. — 447 p.
 152. Heidelberg Institute for International Conflict Research (HIK) (2021): Conflict Barometer 2020, Heidelberg. [Електронний ресурс] – Режим доступу: https://hiik.de/wp-content/uploads/2021/03/ConflictBarometer_2020_1.pdf. – Назва з екрану.
 153. Intellectual Analysis of Systemic World Conflicts and Global Forecast for the 21st Century / Mikhael Z. Zgurovsky, YuriyP. Zaychenko // Big Data: Conceptual Analysis and Applications. – Cham, Switzerland: Springer International Publishing, 2020. – Chapter 4. – P. 141–277. – (Studies in Big Data; Series ISSN2197-6503 ; Volume58).
 154. Розробка методики визначення рівня загроз сталому економічному розвитку України / М. З. Згуровський, О. А. Гавриш, С. В. Войтко // Економічний вісник Національного технічного університету України «Київський політехнічний інститут». – 2011. – № 8. – С. 26–33.

ПРИКЛАД МОДЕЛЮВАННЯ ПРОЦЕСІВ СТАЛОГО РОЗВИТКУ НА ПЛАТФОРМІ МІЖДИСЦИПЛІНАРНИХ ДОСЛІДЖЕНЬ СЦД-УКРАЇНА

У підрозділі 4.3 розглянуто модель кількісного оцінювання процесів сталого розвитку в контексті якості та безпеки життя людей, введено необхідні метрики для його вимірювання.

У цьому додатку подано процес організації і проведення міждисциплінарних досліджень на базі СЦД-Україна з комплексного моделювання процесів сталого розвитку країн світу і регіонів України.

Аналіз процесів сталого розвитку країн світу і регіонів України, а також оцінювання впливу сукупності глобальних погроз на сталий розвиток цих об'єктів є класичним прикладом складної міждисциплінарної задачі, що вимагає вибірки й обробки великих масивів геофізичних і соціально-економічних даних з різнотипних джерел, що суттєво відрізняються одне від одного за форматами подання й зберігання даних.

Розглянемо роботу інструментарію, реалізованого на базі розроблених платформи підтримки міждисциплінарних досліджень, для он-лайн моделювання і порівняльного аналізу соціально-економічного розвитку України в глобальному розрізі (в розрізі країн Європи та Світу), а також аналізу поточного стану соціально-економічного розвитку окремих регіонів (областей) України з позицій концепції сталого розвитку.

Цей інструментарій дозволяє досліджувати динаміку індексу сталого розвитку та його складових, здійснювати порівняльний аналіз досліджуваних об'єктів (адміністративно-територіальних одиниць) за широким спектром показників сталого розвитку (рис. 4.17, 4.18, табл. 4.6, 4.7).

Так, на рис. А.1 представлено загальний вигляд тематичної панелі для аналізу динаміки компоненти якості життя людей країн Європи.

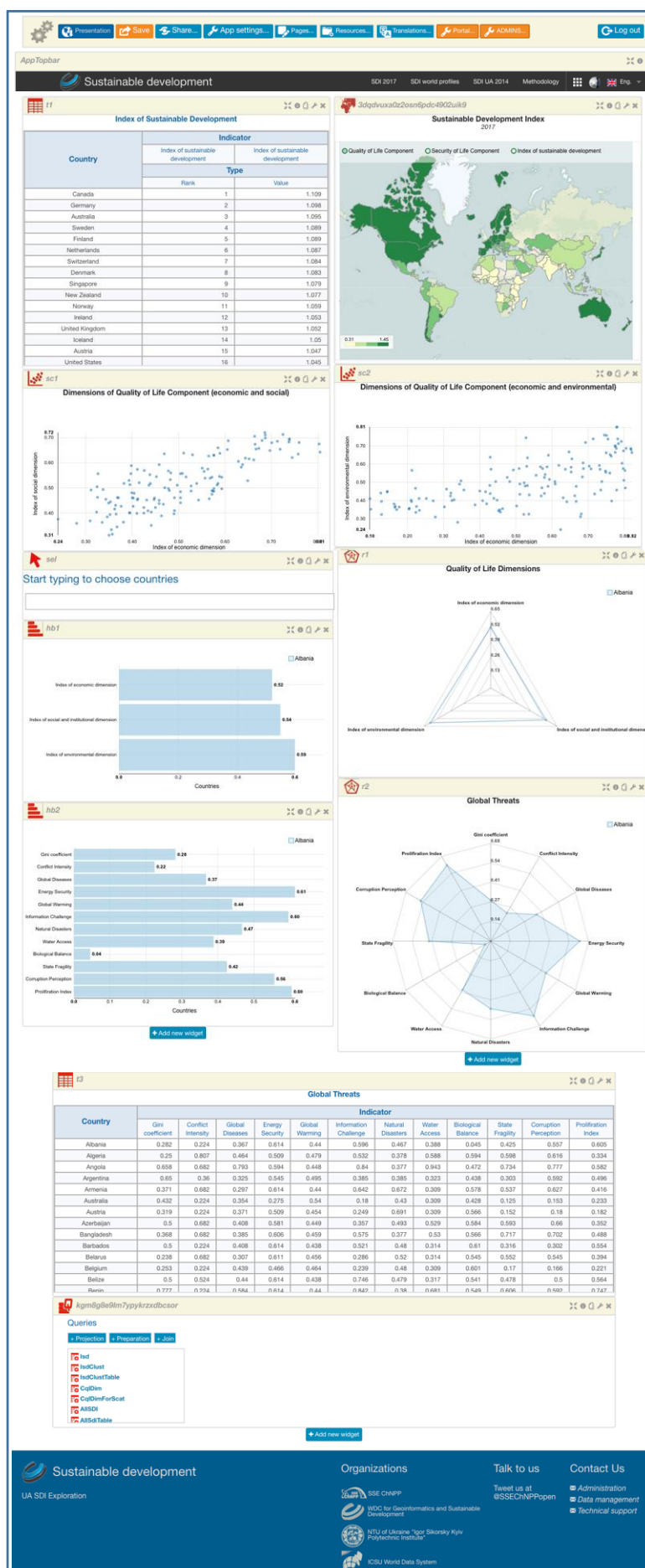


Рис. А.1. Компонента якості життя людей для країн Європи

що містить інструменти для аналітичного опрацювання та візуалізації відповідних даних.

На рис. А.2 у вигляді карти та в табл. А.1 представлено результати роботи інструменту кластеризації країн Європи за Індексом сталого розвитку.



Рис. А.2. Порівняння країн Європи за Індексом сталого розвитку

Таблиця А.1.

Країни, вибрані для порівняння з Україною за Індексом сталого розвитку

Країна	2014 рік	
	Значення індексу	Група
Азербайджан	0.7	2
Молдова	0.71	2
Україна	0.71	2
РФ	0.71	2
Боснія і Герцеговина	0.72	2

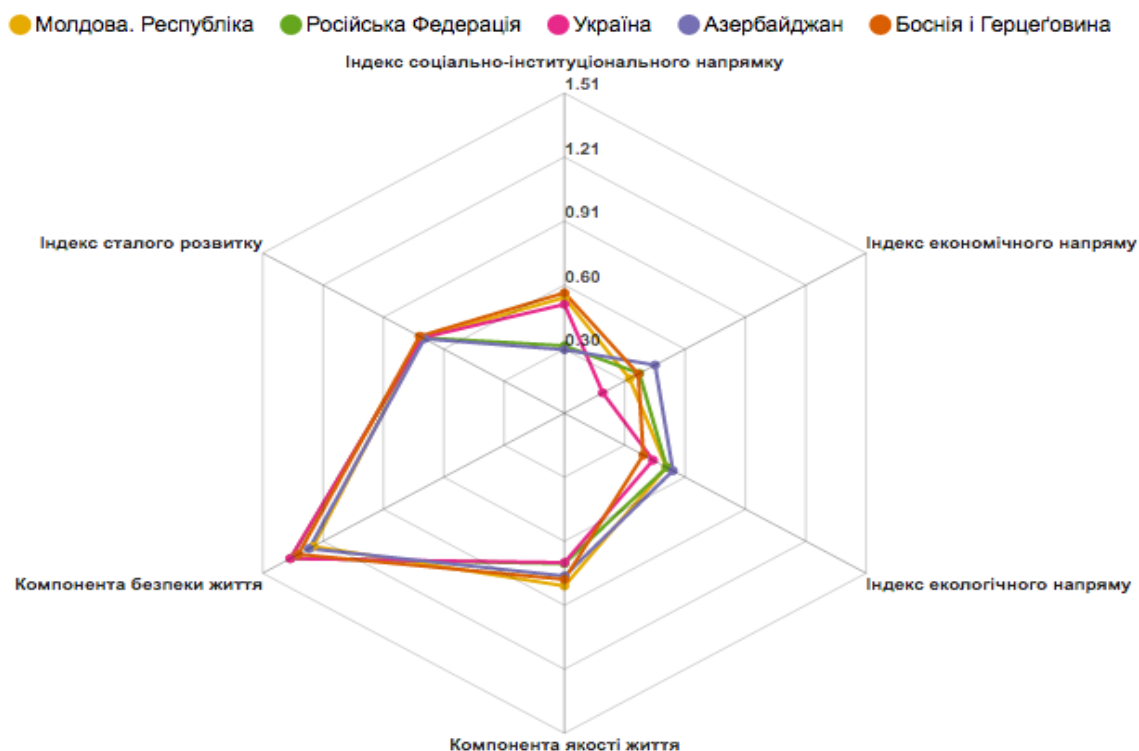


Рис. А.3. Показники сталого розвитку для країн з групи 2

Бачимо, що Україна входить до групи країн зі значеннями Індексу сталого розвитку близькими до 0.7, але, як видно з пелюсткової діаграми, представленої на рис. А.3, розбіжність значень вимірів (соціальний, економічний, екологічний) компоненти якості життя людей є значною, особливо за економічним виміром. Дослідження кореляції між значеннями Індексу сталого розвитку для країн Європи (рис. А.4) свідчить про те, що Україна має динаміку соціально-економічного розвитку, яка суттєво відрізняється від динаміки країн Європи. В таблиці 4.4 наведено групу країн, близьких за значеннями вимірів Компоненти якості життя людей.

З рис. А.5 випливає, що, хоча країни з групи порівняння з близькими значеннями Індексу Компоненти якості життя людей (таблиця А.2) мають схожу структуру Компоненти якості життя людей, проте суттєво відрізняються за Індексом економічного виміру.

Як видно з рис. А.6, Україна за динамікою змін Компоненти якості життя людей суттєво відрізняється від інших країн з зазначеної групи порівняння. Найбільш близькою до неї є Молдова.

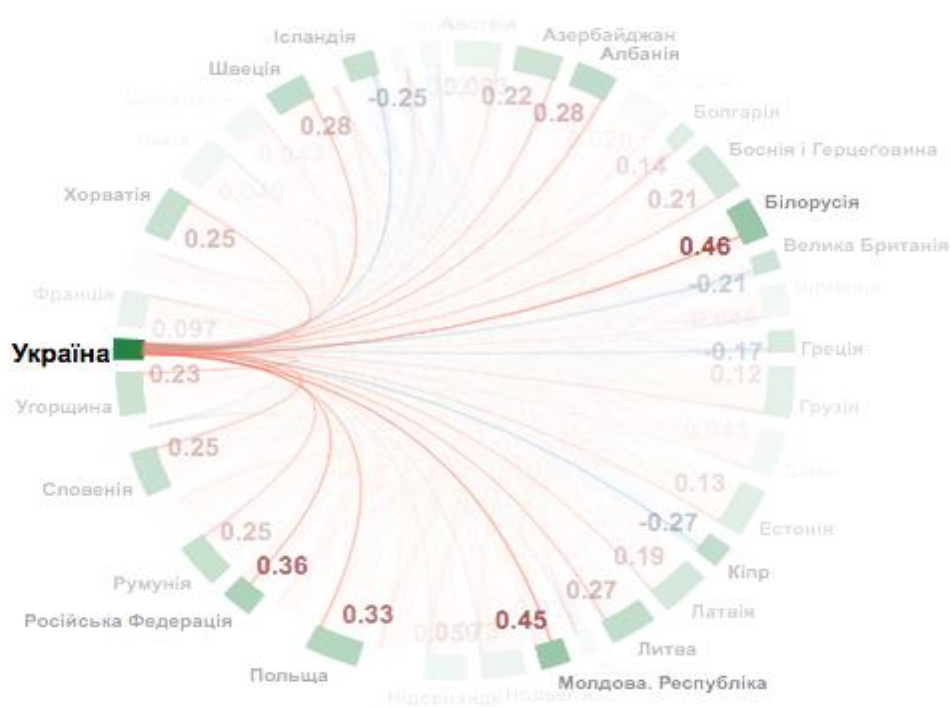


Рис. А.4. Кореляція значень Індексу сталого розвитку України і інших країн Європи

Таблиця А.2.

Країни вибрані для порівняння з Україною за складовими компоненти якості життя людей

Країна	2014			
	Індекс соціально-інституційного напрямку	Індекс економічного напрямку	Індекс екологічного напрямку	Група
Молдова	0.55	0.33	0.51	1
Румунія	0.58	0.56	0.47	1
Грузія	0.53	0.7	0.42	1
Туреччина	0.54	0.54	0.53	1
Україна	0.51	0.19	0.44	1
Боснія і Герцеговина	0.57	0.37	0.39	1

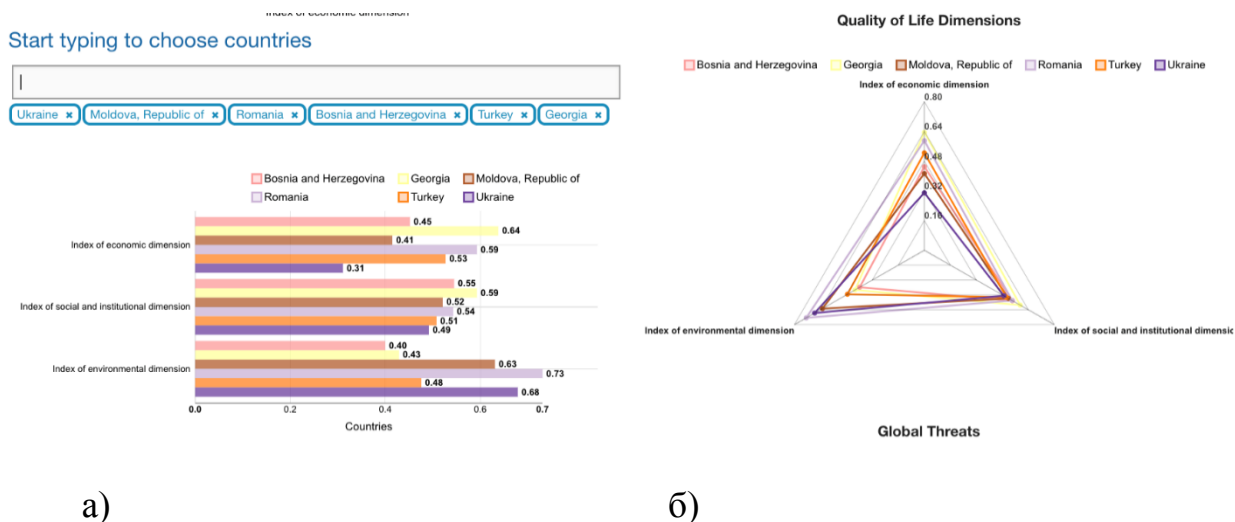


Рис. А.5. Значення складових Компоненти якості життя людей для країн з групи порівняння (Таблиця А.2): а) стовпчикова, б) пелюсткова діаграми

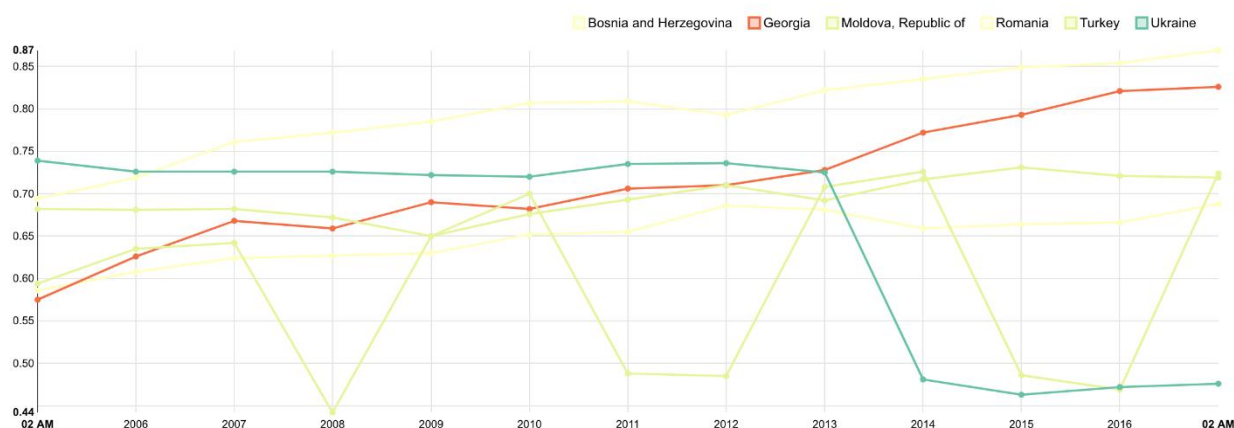


Рис. А.6. Динаміка Компоненти якості життя людей для країн з групи порівняння (Таблиця А.2)

Аналогічно до випадків, представлених на рис. А.6 можна розглянути групу країн, до якої входить Україна, за значенням Компоненти безпеки життя людей. До цього кластеру відносяться Туреччина, Кіпр, Греція, Грузія, Російська Федерація, Болгарія, Молдова, Україна, Боснія і Герцеговина, Білорусія, Албанія, Вірменія, Азербайджан (Рис. А.7).

Профілі загроз для цієї групи зображено на рис. А.8. Бачимо, що за вразливістю до загроз Україна найбільш наближена до Азербайджану, Вірменії та Молдови.

Проведене дослідження показує, що з країн сусідів України, за значною кількістю індикаторів сталого розвитку, найбільш наближеною є Молдова.



Рис. А.7. Групи порівнянь країн Європи за Компонентою безпеки

Global Threats

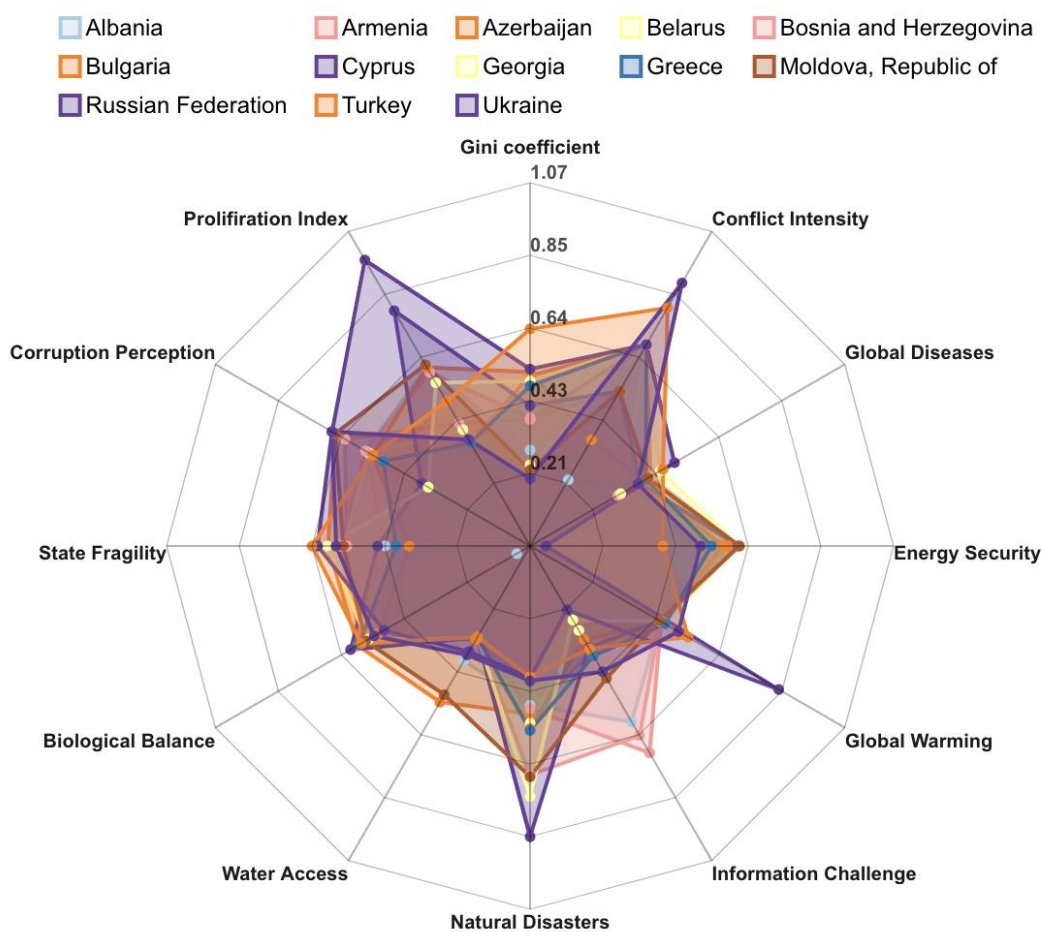


Рис. А.8. Профілі загроз

Необхідно зазначити, що навіть в своїй групі країн за Індексом сталого розвитку Україна показує низькі значення економічного виміру

Таким чином, за допомогою запропонованого інструментарію можна оцінювати наближення (чи віддалення) України до «Оптимістичного» сценарію та сценарію «Збалансований розвиток» [A7]. Інструментарій дозволяє також провести більш детальний аналіз показників країн та їх регіонів, за окремими індикаторами, провести оцінювання стратегій та порівняння сценаріїв розвитку країн в динаміці тощо.

Застосування інструментарію для моделювання та порівняльного аналізу показників сталого розвитку в розрізі регіонів (областей) України. По-суті виконуються такі ж операції, як для глобального аналізу, і використовуються аналогічні інструменти, але опрацьовуються інші дані.

На рис. А.9 наведено загальний вигляд тематичної панелі для аналізу поточного стану соціально-економічного розвитку регіонів (областей) України з позицій концепції сталого розвитку. Панель містить інструменти, які дозволяють:

- визначати групи регіонів (областей), близьких за значеннями індикаторів сталого розвитку;
- здійснювати порівняльний аналіз регіонів (областей) України за Індексом сталого розвитку (компонентами безпеки та якості життя людей і їх складовими);
- знаходити групи регіонів (областей) за схожою структурою загроз та здійснювати порівняльний аналіз за масштабом їх проявів.

Це дозволяє поділити регіони України за Індексом сталого розвитку, Індексом якості і безпеки життя людей на групи: з високим рівнем значення обох індексів; з високим значенням Компоненти безпеки життя та відносно низьким рівнем економічного розвитку; з середнім значенням обох індексів; низькими значеннями обох індексів.

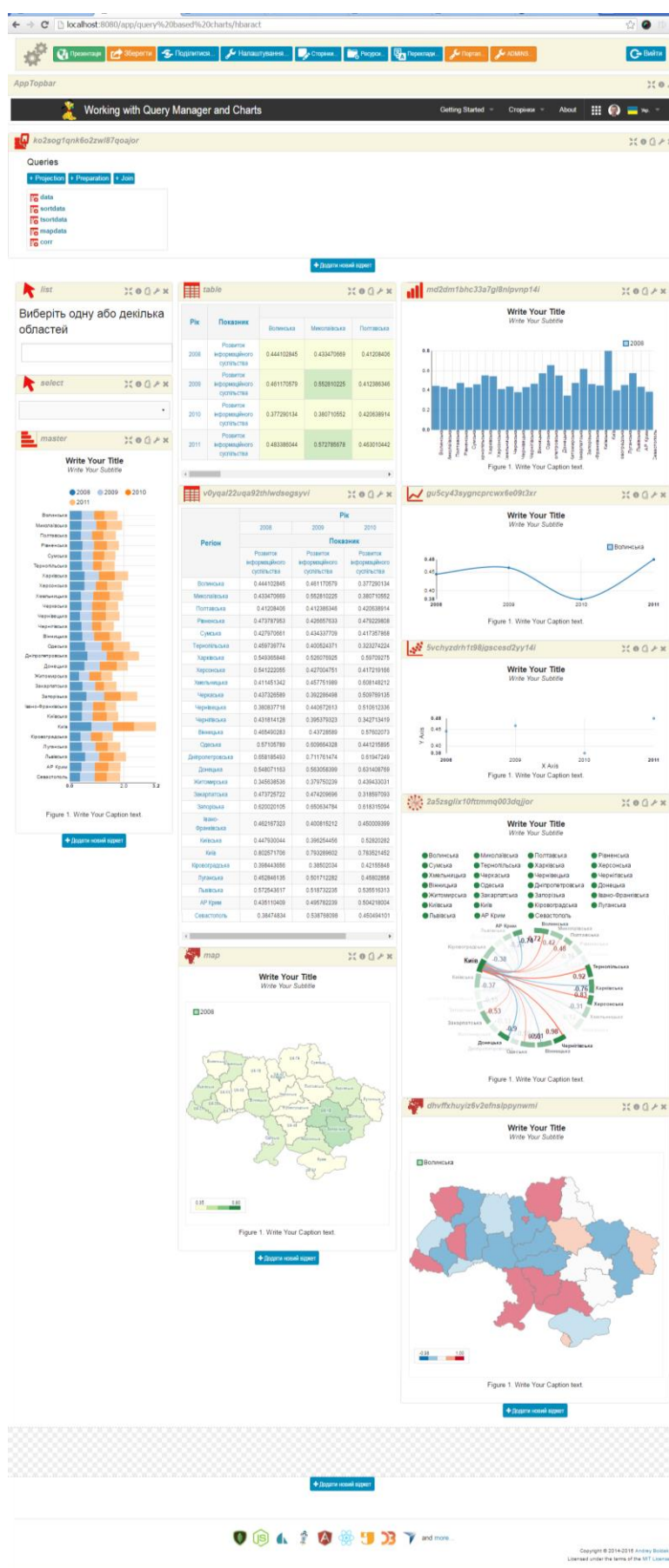


Рис. А.9. Тематична панель для аналізу поточного стану соціально-економічного розвитку регіонів (областей) України

**ОЦІНЮВАННЯ СТАВЛЕННЯ СУСПІЛЬСТВА ДО ПРОВЕДЕННЯ В
УКРАЇНІ МАСОВОЇ ВАКЦИНАЦІЇ ПРОТИ COVID-19,
ВИКОНАНЕ НА ПЛАТФОРМІ МІЖДИСЦИПЛІНАРНИХ
ДОСЛІДЖЕНЬ СЦД-УКРАЇНА**

У підрозділі 4.3 розглянуто методику оцінювання ефективності суспільних перетворень, в основі якої лежить визначення міри неузгодженості між діями влади і очікуваннями суспільства та синергії (соціальної активності) людей, в залежності від вищезазначеної міри неузгодженості.

У цьому додатку подано процес організації і проведення досліджень з оцінювання ставлення суспільства до щеплення проти COVID-19 на основі аналізу емоційного забарвлення інформаційних повідомлень, отриманих шляхом онлайн-моніторингу медіа-ресурсів та соціальних мереж за допомогою інструментарію для оцінювання ефективності масових суспільних перетворень «PRO ET CONTRA» інтегрованої онлайн-платформи «Advanced Analytics» Світового центру даних з геїонформатики та сталого розвитку», який працює на базі КПІ ім. Ігоря Сікорського. Платформу «Advanced Analytics» СЦД-Україна створено на основі розробленої в рамках дисертаційної роботи платформи підтримки міждисциплінарних досліджень.

Аналіз виконано за період спостереження з 01.07.2020 року по 08.01.2021 рік на наборах даних двох інформаційних систем:

1) Системи моніторингу соціальних медіа та лінгвістичного аналізу великих обсягів даних Robusta Світового центру даних «Геїонформатика та сталий розвиток» (<http://open.wdc.org.ua/>, далі за текстом – IC Robusta), що охоплює 11 найбільш популярних соціальних медіа та 500 новинних українських веб-сайтів, соціальних мереж і веб-ресурсів загальної тематики. Обсяг вибірки становить понад 3 млн. повідомлень;

2) Системи автоматизованого збору новинної інформації з відкритих джерел мережі Інтернет, її обробки, систематизації та доступу до неї InfoStream (<http://infostream.ua/>, далі за текстом – IC InfoStream), що охоплює близько 5000 джерел – усі основні інформаційні сайти України, а також провідні закордонні Інтернет-ресурси.

Запит 1. Аналіз інформаційних повідомлень стосовно вакцинації проти COVID-19 в Україні

За результатами аналізу повідомлень, які містять у своєму тексті поєднання фраз «вакцина», «щеплення», «COVID», «коронавірус», «коронавирус», «SARS-COV-2», «Украина», «Україна» в різних відмінках і множині (текст запиту мовою IC Robusta (тут і надалі реєстр літер не має значення, враховуються всі морфологічні форми слів): *(вакцина|щеплення)&(COVID|коронавірус|коронавирус|SARS-COV-2)&(украина|україна)*), далі за текстом – Запит 1) за період з 01.07.2020 року по 08.01.2021 в системі Robusta було відібрано 12 113 повідомлень, з яких 23% носять негативний характер, 24% мають позитивне емоційне забарвлення і 53% є нейтральними (серед них і такі, для визначення емоційного забарвлення яких не достатньо інформації).

На основі цих даних розраховано оцінки норми вектора очікувань суспільства $|\vec{S}| = 0.530$ та кут $\alpha = 89^\circ$. На векторній діаграмі в центрі тематичної панелі (рис. Б.1.) зображено співвідношення між векторами можливих дій влади \vec{G} , очікувань суспільства \vec{S} та показником суспільних перетворень $R=1,03$. Коефіцієнт конфліктності $C = 0,47$.

За ідентичним запитом з системи InfoStream було отримано 96 449 повідомлень, з них 9% мають позитивне емоційне забарвлення, 10% – негативне, 81% – нейтральне (в т.ч. невизначені). Норма вектора очікувань суспільства $|\vec{S}| = 0.81$, кут між векторами дій влади \vec{G} та соціальної активності (очікувань) суспільства \vec{S} $\alpha = 90,5^\circ$, показник суспільних перетворень $R = 0,98$, коефіцієнт конфліктності $C = 0,19$.



Рис. Б.1. Результати оцінювання ставлення суспільства до щеплення проти COVID-19 (за даними IC Robusta, Запит 1), представлені за допомогою тематичних панелей застосунку «PRO ET CONTRA»

Отримані результати свідчать про те, що ставлення українського суспільства до щеплення проти COVID-19 (в інформаційному сегменті проведеного моніторингу) є в значній мірі невизначеним (рівень невизначеності коливається в діапазоні 53%-81%; узагальнений коефіцієнт конфліктності \tilde{C} становить 0,33), при практичному балансі між кількістю позитивних і негативних повідомлень на рівні 10%-24% та незначній позитивній синергії суспільства (узагальнене значення показника суспільних перетворень $\tilde{R} = 1$).

На рис. Б.2 наведено графіки, що відображають добові та інтегровані значення кількості повідомлень з досліджуваної тематики (за даними Robusta). Можемо бачити, що з середини вересня відбувся злам: кількість повідомлень, що мають позитивне емоційне забарвлення стосовно можливого щеплення проти COVID-19, почала переважати кількість негативних повідомлень. Помітні піки спостерігаються з 8 по 22 жовтня 2020 року, з 12 по 20 листопада, 22 і 30 грудня 2020 року, 6 січня 2021 року.

Аналогічну ситуацію ми бачимо і за даними IC InfoStream (рис. Б.3).

Динаміка соціальної активності суспільства

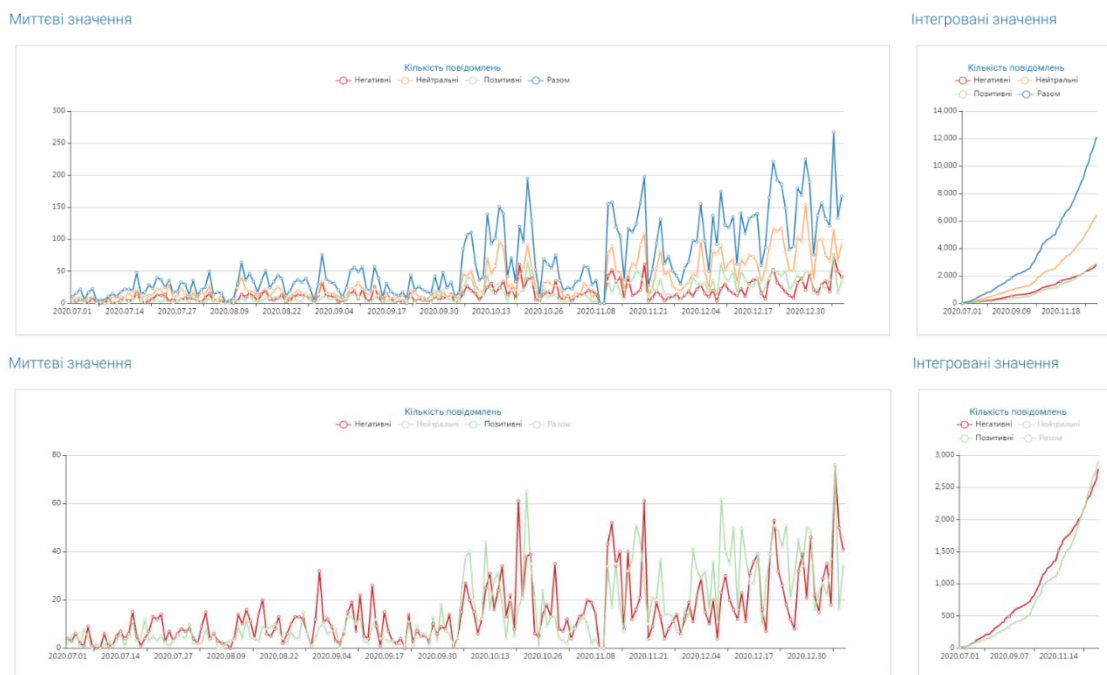


Рис. Б.2. Динаміка суспільної активності (за даними IC Robusta, Запит 1)

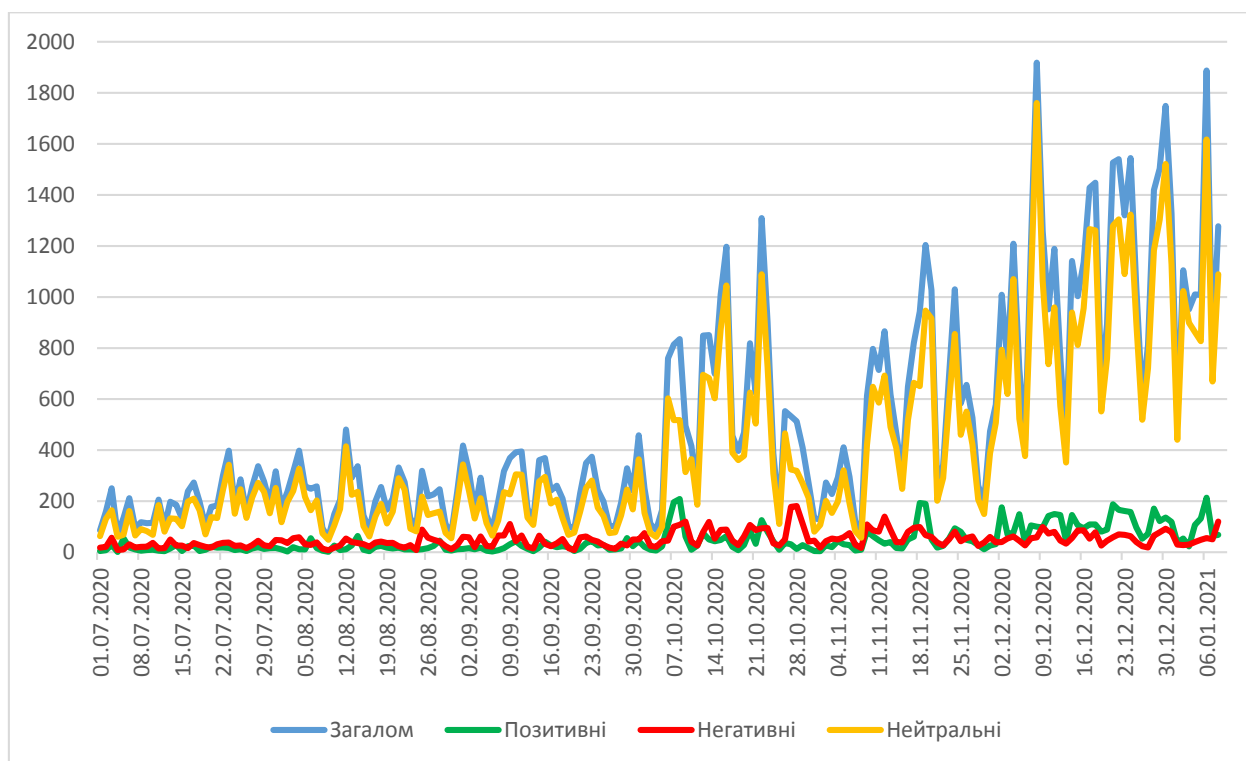


Рис. Б.3. Динаміка суспільної активності (за даними IC InfoStream, Запит 1)

Запит 2. Деталізація запиту 1 з включенням до нього переліку конкретних вакцин

До запиту 1 включається перелік конкретних вакцин від COVID-19: Oxford Uni – AstraZeneca, Moderna, Pfizer – BioNTech, Sinovac Biotech, Gamaleya Sputnik V. Текст запиту мовою Robusta): (вакцина|щеплення) &(AstraZeneca |Pfizer|Moderna|Sinovac|Спутник|Сунутник|Sputnik)& (украина|україна), далі за текстом – Запит 2.

За запитом в системі Robusta було відібрано 3486 повідомлень, з них 17% негативних та 29% позитивних; в системі Infostream – 30 210 повідомлень, з яких 6% негативних і 8% позитивних. Відповідно, за даними Robusta норма вектора очікувань суспільства $|\vec{S}| = 0.547$, кут $\alpha = 77^\circ$, показник суспільних перетворень $R = 1.39$, коефіцієнт конфліктності C становить 0,44 (рис. Б.4).

За даними InfoStream $|\vec{S}| = 0.863$, кут $\alpha = 88^\circ$, показник суспільних перетворень $R = 1.08$, коефіцієнт конфліктності $C = 0.13$. Графіки добової кількості повідомлень (див. рис. Б.5, Б.6) мають пікові значення у першій декаді, кінці грудня 2020 р. та початку січня 2021 року (06.01.2021 р.).

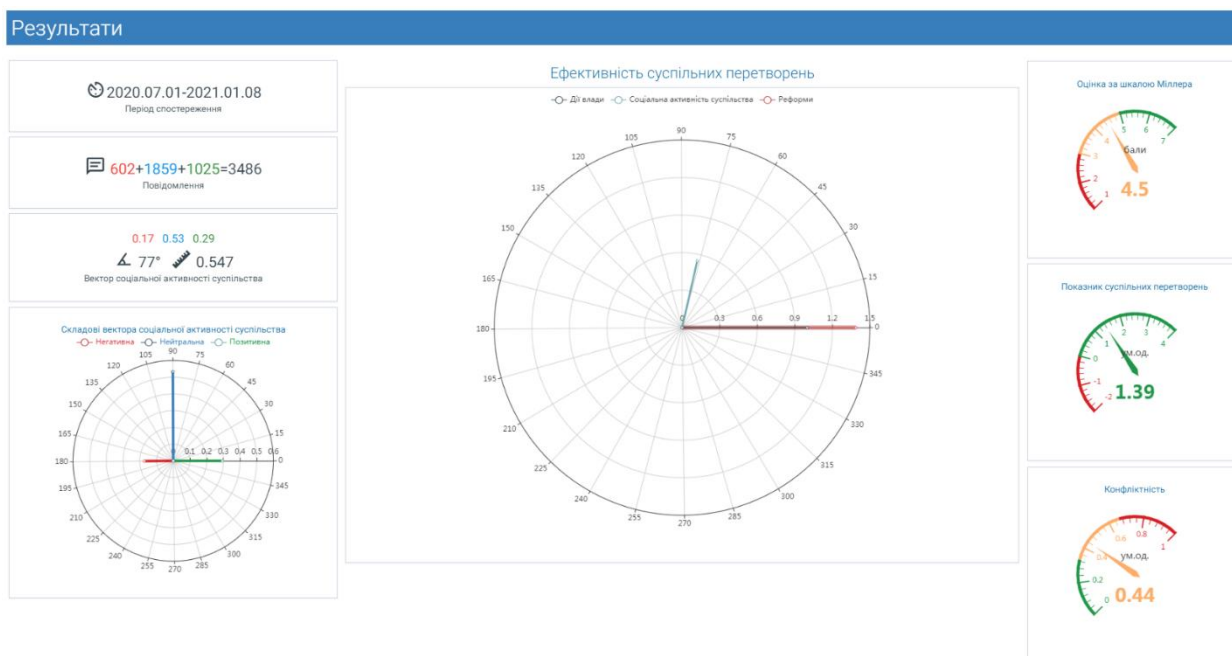


Рис. Б.4. Результати оцінювання ставлення суспільства до щеплення проти COVID-19 (за даними IC Robusta, Запит 2)

Результати Запиту 2 показують незначну перевагу позитивних очікувань (над негативними і нейтральними) щодо можливого щеплення проти COVID-19 (має місце синергетичний ефект – узагальнене значення показника суспільних перетворень \tilde{K} становить 1,20 і перевищує значення норми вектора дій влади \vec{G}), узагальнений коефіцієнт конфліктності $\tilde{C} = 0,29$.

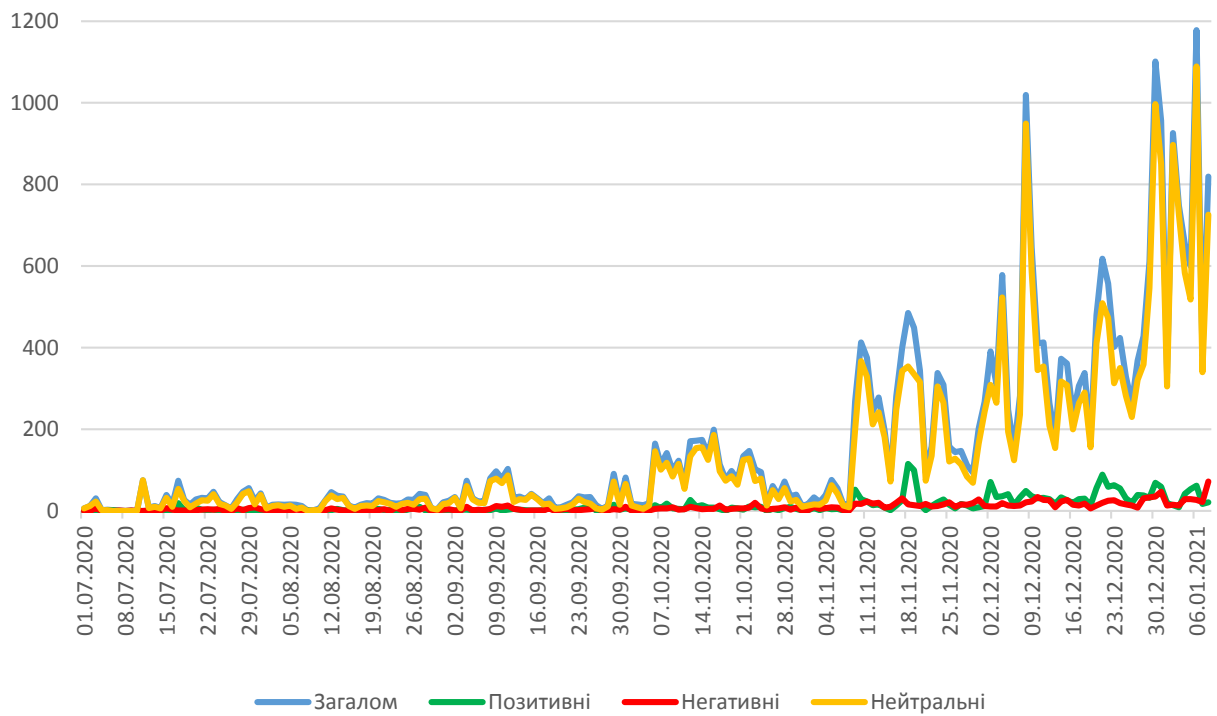


Рис. Б.5. Динаміка суспільної активності (за даними InfoStream, Запит 2)

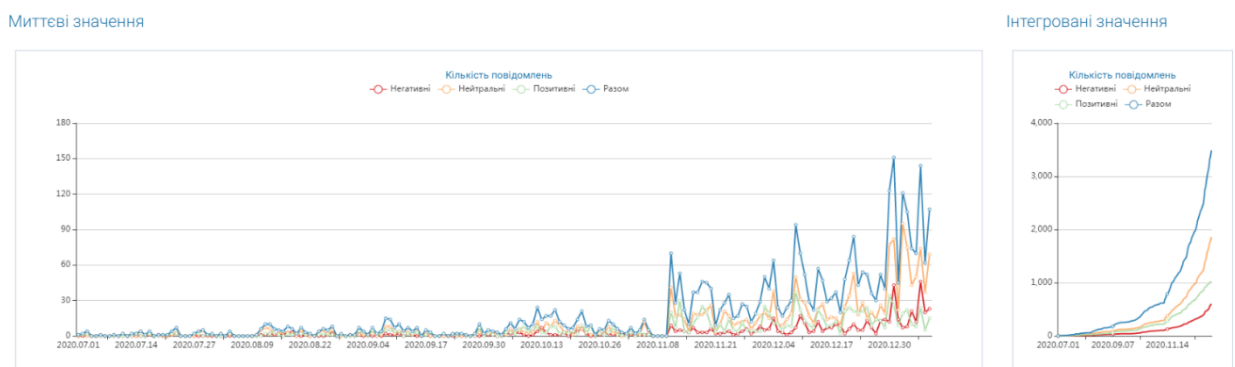


Рис. Б.6. Динаміка суспільної активності (за даними Robusta, Запит 2)

Запит 3. Ставлення суспільства до щеплення проти COVID-19 окремо кожною з вакцин

Проаналізуємо ставлення суспільства до щеплення проти COVID-19 окремо кожною з вакцин: *Oxford Uni* – *AstraZeneca*, *Moderna*, *Pfizer* –

BioNTech, Sinovac Biotech, Gamaleya Sputnik V (далі за текстом – Серія запитів 3). Результати аналізу наведено в таблиці А.1.

Відповідно до поточного рівня інформованості населення України за узагальненим показником суспільних перетворень вакцини розташовані у такому порядку: перше місце і найвищий рівень довіри має вакцина Oxford Uni – AstraZeneca ($\tilde{R} = 1,49$); на другому місці вакцина Moderna ($\tilde{R}=1,45$); на третьому місці вакцина Sinovac Biotech ($\tilde{R}=1,32$); на четвертому місці вакцина Pfizer – BioNTech ($\tilde{R}=1,26$); на останньому місці вакцина Gamaleya Sputnik V ($\tilde{R}=1,13$).

Запит 4. Ставлення українського суспільства до реєстрації і застосування в Україні вакцини Gamaleya Sputnik V

Високий рівень суспільної дискусії стосовно можливої реєстрації і застосування в Україні вакцини Gamaleya Sputnik V обумовив виконання окреме дослідження щодо ставлення українського суспільства до цього явища. Текст запиту мовою Robusta: *(регистрация|реєстрація)&(Спутник|Супутник|Sputnik)&(украина|україна)*, далі за текстом – Запит 4.

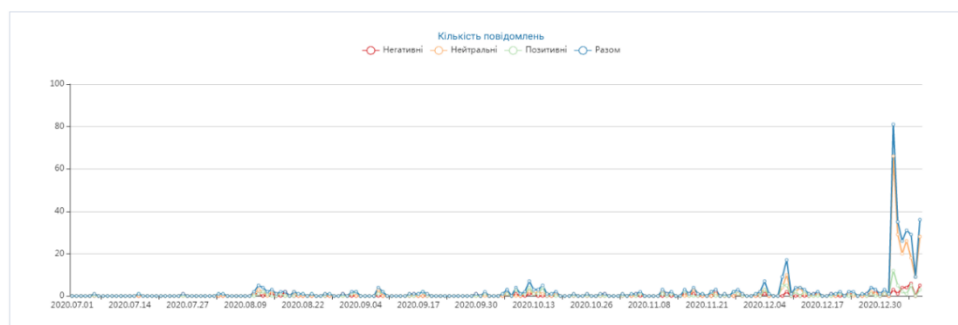
За запитом в системі Robusta відібрано 445 повідомлень (14% негативних, 24% позитивних, 62% нейтральних), в системі Infostream – 3702 повідомлення (6% – негативних, 8% – позитивних, 86% – нейтральних). За даними Robusta норма вектора очікувань суспільства $|\vec{S}| = 0.633$, кут $\alpha = 81^\circ$, показник суспільних перетворень $R = 1,31$, коефіцієнт конфліктності $C = 0,36$. За даними InfoStream: $|\vec{S}| = 0.857$; $\alpha = 89^\circ$; $R = 1,04$; $C = 0,14$. Обчислення на основі даних IC Robusta і IC Infostream дали узагальнене значення показника суспільних перетворень $\tilde{R} = 1,16$ і узагальнений коефіцієнт конфліктності $\tilde{C} = 0,25$.

Аналізуючи динаміку повідомлень стосовно можливої реєстрації і застосування в Україні вакцини Gamaleya Sputnik V (рис. Б.7, Б.8), можемо спостерігати два значні сплески: 09.12.2020 та 02.01.2021 року.

Таблиця Б.1. Результати моніторингу ставлення суспільства до щеплень проти COVID-19 окремими вакцинами.

Назва вакцини	Рейтинг довіри населення до вакцини	Узагальнений показник суспільних перетворень \tilde{R}	Узагальнений коефіцієнт конфліктності \tilde{C}	Джерело даних	Кількість повідомлень				Норма вектора очікувань суспільства $ \vec{S} $	Кут α	Показник суспільних перетворень R	Коефіцієнт конфліктності C
					Загальна	Позитивні	Нейтральні	Негативні				
Oxford Uni – AstraZeneca	1	1,49	0,29	Robusta	586	42%	45%	13%	0,533	57	1,92	0,44
				InfoStream	6609	11%	85%	4%	0,852	85	1,22	0,13
Moderna	2	1,45	0,27	Robusta	686	37%	51%	12%	0,566	64	1,79	0,40
				InfoStream	7471	12%	84%	4%	0,844	85	1,22	0,13
Sinovac Biotech	3	1,32	0,21	Robusta	361	32%	57%	11%	0,608	70	1,65	0,35
				InfoStream	3883	5%	92%	2%	0,923	88	1,09	0,07
Pfizer – BioNTech	4	1,26	0,31	Robusta	1298	34%	48%	18%	0,503	72	1,49	0,49
				InfoStream	13905	9%	86%	5%	0,864	87	1,13	0,13
Gamaleya Sputnik V	5	1,13	0,31	Robusta	1261	29%	52%	19%	0,525	80	1,31	0,47
				InfoStream	11575	8%	85%	7%	0,848	90	1,02	0,15

Миттєві значення



Інтегровані значення

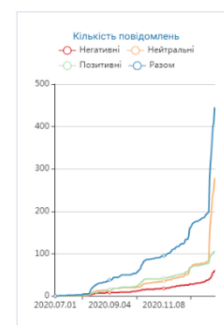


Рис. Б.7. Добова динаміка повідомлень стосовно реєстрації і застосування в Україні вакцини Gamaleya Sputnik V (Запит 4, IC Robusta)

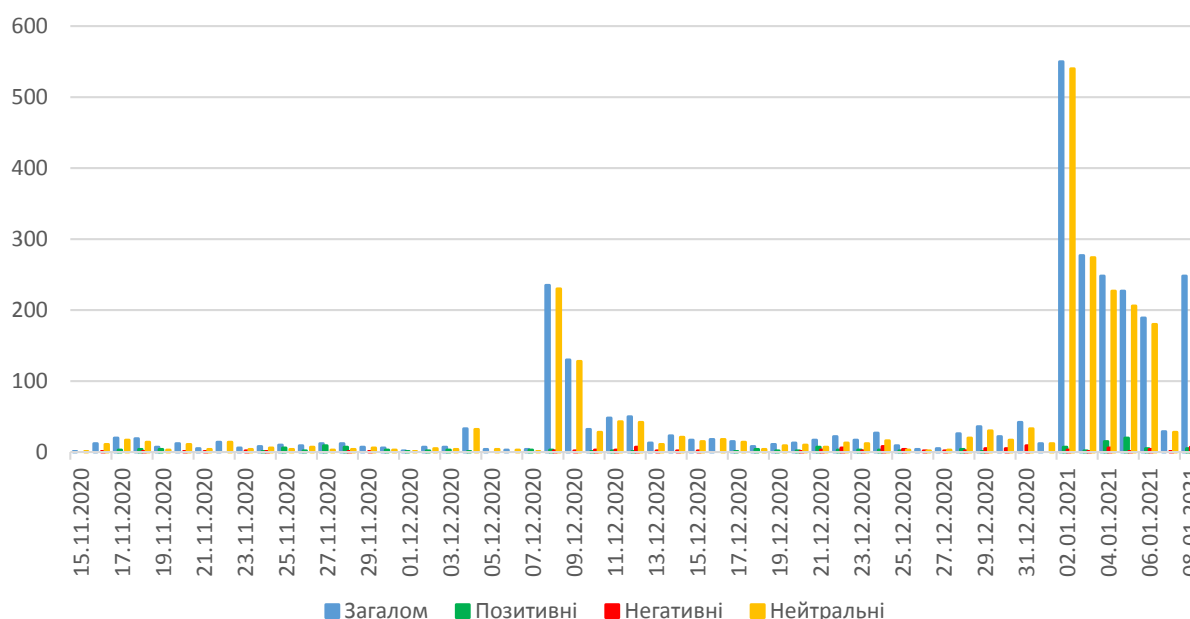


Рис. Б.8. Добова динаміка повідомлень (Запит 4, IC Infostream)

Отже проведені дослідження на основі моніторингу IC Robusta вказує на 14% негативних відгуків, 24% позитивних і 62% нейтральних, в системі Infostream – 6% негативних відгуків, 8% позитивних, 86% нейтральних. Отримані результати свідчать про високий рівень невизначеності ставлення українського суспільства до доцільності реєстрації і застосування в Україні вакцини Gamaleya Sputnik V (від 62% до 86%), з незначною перевагою позитивного ставлення (8%-24%) по відношенню до негативного (6%-14%). Реєстрація і застосування в Україні вакцини Gamaleya Sputnik V можуть бути здійснені з показником суспільних перетворень $\tilde{R} = 1,16$, коефіцієнтом поляризації суспільних думок (коефіцієнтом конфліктності) $\tilde{C} = 0,25$.

Висновки

Аналіз ставлення українського суспільства до щеплень проти COVID-19, виконані за допомогою інтегрованої онлайн-платформи «Advanced Analytics» Світового центру даних «Геіонформатика та сталий розвиток» КПП ім. Ігоря Сікорського на наборах даних двох інформаційних систем: IC Robusta (СЦД-Україна) та IC InfoStream (ІЦ "Електронні вісті"), дозволив отримати обґрунтовані результати, які можна враховувати при прийнятті рішень:

1. Від 53% до 81% інформаційних повідомлень (запити 1, 2 в IC Robusta та InfoStream) не мають яскраво вираженого позитивного чи негативного емоційного забарвлення, що свідчить про невизначеність ставлення значної частини населення України до вакцинації проти COVID-19. Від 9% до 29% інформаційних повідомлень мають позитивне забарвлення, від 10% до 23% – негативне. За умови проведення активної інформаційної політики і ефективних дій влади масова вакцинація населення України проти COVID-19 може бути здійснена з показником суспільних перетворень \tilde{R} від 1,0 до 1,2.

2. Стосовно застосування конкретних вакцин (Запит 3), то найбільша суспільна довіра (при поточному, досить обмеженому інформуванні суспільства про характеристики окремих вакцин) віддається вакцині Oxford Uni – AstraZeneca.

3. Отримані результати вказують на високий рівень невизначеності ставлення населення до доцільності реєстрації і застосування в Україні вакцини Gamaleya Sputnik V (від 62% до 86%), з незначною перевагою позитивного ставлення (8%-24%) по відношенню до негативного (6%-14%).

4. Підготовка до масової вакцинації населення України проти COVID-19 та її проведення повинні підтримуватися активним інформаційним супроводом на державному рівні. З метою планування раціональної масової вакцинації доцільним є проведення оцінки дійсної захворюваності населення України на COVID-19 шляхом створення мережі збору інформації щодо виявлення випадків COVID-19, які не підтверджувалися методом ПЛР та не були включені до офіційної статистики.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ, ВІДОМОСТІ ПРО АПРОБАЦІЮ РЕЗУЛЬТАТІВ ТА ДОКУМЕНТИ, ЯКІ ПІДТВЕРДЖУЮТЬ РЕЗУЛЬТАТИ ВПРОВАДЖЕННЯ

Список публікацій здобувача за темою дисертації, відомості про апробацію результатів

Монографії

1. Zgurovsky M., Boldak A., Melnyk O., Perestyuk M., Putrenko V., Pyshnograiev I., Yasinsky V., Yefremov K., Foresight 2018: systemic world conflicts and global forecast for XXI century / International Council for Science etc.; Scientific Supervisor M. Zgurovsky. — К. : NTUU «Igor Sikorsky Kyiv Polytechnic Institute», 2018. — 226 p. *Здобувачем запропоновано та розроблено мікросервіси для аналізу та прогнозування світових конфліктів; реалізацію розроблених мікросервісів імплементовано до інтегрованої он-лайн платформи Advanced Analytics ННК «Світовий центр даних з геоінформатики та сталого розвитку».*
2. Згуровський М.З., Болдак А.О., Джигирей І.М., Єфремов К.В., Форсайт 2018: Аналіз підготовки і перепідготовки фахівців природничого і технічного спрямування, виходячи з цілей сталого соціально-економічного розвитку України до 2025 року. — К. : НТУУ «КПІ імені Ігоря Сікорського», Вид-во «Політехніка», 2018. — 32 с. *Здобувачем створено та реалізовано засобами розробленої платформи підтримки міждисциплінарних досліджень сценаріїв (ланцюг) послідовного та паралельного виконання мікросервісів для виконання аналізу підготовки і перепідготовки фахівців, запропоновано і реалізовано додаткові необхідні мікросервіси в рамках зазначеної платформи. Результати дослідження представлено за допомогою інструментів розробленої платформи.*
3. Foresight and construction of the strategies of socio-economic development of Ukraine on mid-term (up to 2020) and long-term (up to 2030) time horizons / Scientific advisor of the project acad. of NAS of Ukraine M. Zgurovsky // International Council for Science (ICSU); Committee for the System Analysis of the Presidium of NAS of Ukraine; National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»; Institute for Applied System Analysis of MES of Ukraine and NAS of Ukraine; World Data Center for Geoinformatics and Sustainable Development; Agrarian Superstate Foundation. — 2nd ed. — Kyiv : NTUU «Igor Sikorsky KPI», Publ. house «Polytechnica», 2016. — 184 p. *Здобувачем розроблено застосунки на основі мікросервісів платформи підтримки міждисциплінарних досліджень для проведення сценарного моделювання поведінки складних соціально-економічних систем; виконано інтеграцію необхідних джерел даних, створено віджети для відображення результатів моделювання.*
4. Sustainable Development Analysis: Global and Regional Contexts / International Council for Science (ICSU) and others; Scientific Supervisor of the Project M. Zgurovsky. — К. : Igor Sikorsky KPI, 2017. — Part 2. Ukraine in Sustainable Development Indicators (2016–2017). — 72 p. *Здобувачем запропоновано та реалізовано процедури розрахунку індексу сталого розвитку країн світу, компоненти безпеки життя, компоненти якості життя та її складових з використанням логічного підходу для планування взаємодії мікросервісів платформи підтримки міждисциплінарних досліджень. Здобувачем виконано опис та інтеграцію (засобами платформи) інформаційних ресурсів, необхідних для дослідження.*
5. Sustainable Development Analysis: Global and Regional Contexts / International Council for Science (ICSU) and others; Scientific Supervisor of the Project M. Zgurovsky. — К. : Igor Sikorsky KPI, 2017. — Part 1. Global Analysis of Quality and Security of Life (2016). — 208 p. *Здобувачем запропоновано та реалізовано процедури розрахунку індексу сталого розвитку регіонів України, компоненти безпеки життя, компоненти якості життя та її складових з використанням логічного підходу для планування взаємодії мікросервісів платформи підтримки міждисциплінарних досліджень. Здобувачем виконано опис та інтеграцію (засобами платформи) інформаційних ресурсів, необхідних для дослідження.*

Стаття у періодичному науковому виданні іншої держави, яка входить до ОЕСР та Європейського Союзу

6. Marsel Shaimardanov, Alexei Gvishiani, Michael Zgurovsky, Alexander Sterin, Alexander Kuznetsov, Natalia Sergeyeva, Evgeny Kharin, Kostiantyn Yefremov. Development of WDS Russian-Ukrainian segment // Data Science Journal. – 2013. – Volume 12. – p. 17-26. (журнали країн, що **входять до ОЕСР, ЄС**; включене до бази даних **Scopus**). *Здобувачем описано і проаналізовано типовий процес інтелектуальної обробки даних в міждисциплінарних дослідженнях, сформульовано вимоги до моделей і методів інтелектуальної обробки даних в міждисциплінарних дослідженнях, запропоновано архітектурні принципи організації системи інтелектуальної обробки даних в рамках інформаційно-комунікаційної інфраструктури Світового центру даних з геоінформатики та сталого розвитку і його партнерів.*

Статті у наукових фахових виданнях

7. M.Z. Zgurovsky, A.A. Boldak, K.V. Yefremov. Intelligent analysis and the systemic adjustment of scientific data in interdisciplinary research // Cybernetics and Systems Analysis: Volume 49, Issue 4 (2013), pp. 541-552. (журнал **категорії «А»**, включене до бази даних **Scopus**). *Здобувачем виконано вербальний опис задачі узгодження даних в міждисциплінарних дослідженнях, запропоновано формальні моделі задачі узгодження і розроблено метод узгодження даних з врахуванням оцінок інформаційних втрат процедур перетворення даних та оцінок узгодженості даних, отриманих з різних джерел, для конструювання процедури перетворення даних з найменшими інформаційними втратами.*
8. Теленик С.Ф., Амонс О.А., Єфремов К.В., Лиско В.Т. Логічний підхід до інтеграції програмних застосувань підтримки міждисциплінарних наукових досліджень // Наукові вісті НТУУ «КПІ». – 2013. – № 5. – С. 53-72. *Здобувачем в рамках логічного підходу до інтеграції застосунків запропоновано модифікацію формальної логічної системи на основі клаузальної логіки першого порядку для планування взаємодії застосунків, удосконалений з врахуванням дворівневого характеру логічного формалізму метод виведення для формальної логічної системи і механізм побудови схеми виконання запиту користувача на основі виводу.*
9. Теленик С.Ф., Амонс О.А., Єфремов К.В., Жук С.В. Семантична інтеграція різнорідних інформаційних ресурсів // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: Зб. наук. пр. – К.: Век+, – 2013. – №. 58. – с.29-45. *Здобувачем запропоновано загальний підхід до вирішення проблеми інтеграції інформаційних ресурсів Світової системи даних на основі web-орієнтованих технологій, способу організації взаємодії джерел даних з посередником і моделей дескриптивної логіки в термінах OWL, RDF і SPARQL, а також схему оброблення запитів до розподілених джерел даних, моделі і методи формування плану виконання запиту та його оптимізації, формування запитів безпосередньо до джерел даних, алгоритм побудови зв'язаної структури даних із онтологій та підключень до джерел даних.*
10. Згуровский М.З., Болдак А.А., Ефремов К.В., Сергеева Н.А., Забаринская Л.П., Шестопалов И.П., Нисилевич М.В. Применение методов интеллектуального анализа данных для эмпирических исследований взаимосвязи гелио- и геофизических процессов // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: збірник наукових праць. – К.: Век+, – 2013. – №. 58. – с.4-10. *Здобувачем розроблено на основі логічного підходу і реалізовано засобами платформи підтримки міждисциплінарних досліджень сценарій (сукупність взаємопов'язаних процесів) обробки даних для вирішення задачі аналізу взаємозв'язку геліо- і геофізичних процесів.*
11. Болдак А.О., Єфремов К.В. Предметно-орієнтована мова аналітичної обробки даних // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: збірник наукових праць. – К.: Век+, – 2012. – №. 55. – с.67-71. *Здобувачем запропоновано загальний підхід до розроблення і надання користувачам Світової системи даних нових інформаційних ресурсів і проблемно-орієнтовану графічну мову аналітичної обробки даних як основу швидкого розроблення застосунків на платформі підтримки міждисциплінарних досліджень.*
12. M. Z. Zgurovsky, A. D. Gvishiani, K. V. Yefremov and A. M. Pasichny. Integration of the Ukrainian science into the world data system // Cybernetics and Systems Analysis: Volume 46, Issue 2 (2010), pp. 211-219. (журнал **категорії «А»**, включене до бази даних **Scopus**). *Здобувачем виконано аналіз діяльності Світової системи даних і Світового центру даних з геоінформатики та сталого розвитку як об'єкта управління, виявлено і проаналізовано невирішені проблеми підтримки*

міждисциплінарних досліджень в Світовій системі даних, сформульовано принципи формування IT-середовища, здатного забезпечити ефективну підтримку міждисциплінарних досліджень і формування на цій основі нової IT-спільноти.

Матеріали науково-технічних конференцій, що засвідчують апробацію матеріалів дисертації

13. M. Zgurovsky, A. Boldak, D. Lande, K. Yefremov and M. Perestyuk. Predictive Online Analysis of Social Transformations based on the Assessment of Dissimilarities between Government Actions and Society's Expectations. Conference proceedings of 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2020, P. 130-135. (включене до бази даних **Scopus**). *Здобувачем на базі інтегрованої он-лайн платформи Advanced Analytics Світового центру даних з геоінформатики та сталого розвитку розроблено сукупність інструментів та сервісів для кількісного оцінювання параметрів і характеристик суспільних перетворень (на прикладі кількісного оцінювання ставлення населення України до дій влади, пов'язаних з посиленням карантинних заходів, спрямованих на протидію поширенню пандемії COVID-19).*
14. Nowakowski, G., Telenyk, S., Yefremov, K., Khmeliuk, V. The approach to applications integration for world data center interdisciplinary scientific investigations. Proceedings of the 2019 Federated Conference on Computer Science and Information Systems. FedCSIS 2019, P. 539-545. (включене до бази даних **Scopus**). *Здобувачеві належить архітектура системи інтеграції застосунків в Світовій системі даних для підтримки міждисциплінарних досліджень на основі оркестрації як способу організації взаємодії сервісів, механізм якої базується на методах представлення семантики перетворень, виконуваних сервісами, методах виведення в клаузальній логіці і методах відновлення схем виконання запитів користувачів на основі виводу.*
15. Zgurovsky, M., Putrenko, V., Dzhygyrey, I., Boldak, A., Yefremov, K., Pashynska, N., Pyshnograiev, I., Nazarenko, S. Parameterization of Sustainable Development Components Using Nightlight Indicators in Ukraine. Conference proceedings of 2018 IEEE 1st International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2018, P. 1-5. (включене до бази даних **Scopus**). *Здобувачем запропоновано загальний підхід до оцінювання сталого розвитку територій на основі яскравості нічних вогнів; на базі інтегрованої он-лайн платформи Advanced Analytics Світового центру даних з геоінформатики та сталого розвитку розроблено сукупність інструментів та сервісів для встановлення залежностей між супутниковими даними яскравості нічних вогнів та показниками сталого розвитку на прикладі території України.*
16. Zgurovsky, M., Boldak, A., Yefremov, K., Pyshnograiev, I. Modeling and investigating the behavior of complex socio-economic systems. 2017. IEEE 1st Ukraine Conference on Electrical and Computer Engineering, UKRCON 2017 – Proceedings, P. 1113-1116. (включене до бази даних **Scopus**). *Здобувачем запропоновано реалізацію серії методів для оцінки поточного стану складних соціально-економічних систем, прогнозування їх майбутніх станів; розроблено та реалізовано засобами платформи підтримки міждисциплінарних досліджень сценарії (сукупність взаємопов'язаних мікросервісів) обробки даних для дослідження складних соціально-економічних систем; представлено за допомогою інструментів розробленої платформи результати дослідження.*
17. Telenyk, S., Nowakowski, G., Yefremov, K., Khmeliuk, V. Logics based application integration for interdisciplinary scientific investigations. Proceedings of the 2017 IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2017, P. 1026-1031. (включене до бази даних **Scopus**). *Здобувачеві належить підхід до інтеграції сервісів програмних систем Світової системи даних на основі автоматизованої побудови схеми взаємодії сервісів для виконання запиту користувача шляхом виведення в клаузальній логіці, логічні моделі опису сервісів (застосунків) та джерел даних, методи виведення та відновлення схеми виконання запиту на основі виводу, вимоги до складу і взаємодії компонентів підтримки міждисциплінарних досліджень в Світовій системі даних з урахуванням стану IT-галузі.*

Документи, які підтверджують результати впровадження

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК
УКРАЇНИ
ІНСТИТУТ ГЕОФІЗИКИ
ім. С. І.Субботіна

Україна 03680, Київ, пр. Академіка Палладіна, 32
Тел.: +380 (44) 424-01-12; Факс.: +380 (44) 450-25-20
E-mail: earth@igph.kiev.ua; www.igph.kiev.ua



NATIONAL ACADEMY OF SCIENCES
OF THE UKRAINE
INSTITUTE OF GEOPHYSICS
by S.I.Subbotinname

Academician Palladin, av. 32, Kiev, 03680, Ukraine
Tel.: +380 (44) 424-01-12; Fax.: +380 (44) 450-25-20
E-mail: earth@igph.kiev.ua; www.igph.kiev.ua

01.02.2/р. № 01-8-53
На № _____ від _____

А К Т

про впровадження результатів дисертаційної роботи
Єфремова Костянтина Вікторовича на тему «Платформа підтримки
міждисциплінарних досліджень в Світовій системі даних», поданої на
здобуття наукового ступеня кандидата технічних наук за спеціальністю
05.13.06 – Інформаційні технології

Платформа підтримки міждисциплінарних досліджень в Світовій системі даних та її окремі компоненти, розроблені в дисертаційній роботі Єфремова К. В., використані при створенні інформаційних систем Інституту геофізики ім. С.І. Субботіна Національної академії наук України.

1. Використання запропонованих в дисертаційній роботі моделей і методів та розроблених на їх основі інструментів генерування проблемно-орієнтованих застосунків для аналітичного оброблення міждисциплінарних даних дозволило скоротити час проектування і реалізації тематичних застосунків на 30 %.

2. Використання комплексу моделей і методів інтеграції застосунків і джерел даних дозволило підвищити ефективність аналізу даних різної природи експертами без знання інструментальних мов програмування на 42%.

3. Впровадження запропонованого в роботі методу узгодження даних дозволило зменшити на 7,25% інформаційні втрати при перетворенні даних з різних джерел, представлених в різних шкалах.

4. Застосування архітектурних принципів організації системи інтелектуальної обробки даних, запропонованих в роботі, дозволило скоротити втрати внаслідок невчасного отримання потрібних даних на 11,75 %.

Директор,
академік НАН України, д.ф.-м.н., проф



В. І. Старостенко
В. І. Старостенко



НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ІНСТИТУТ ПРОБЛЕМ РЕЄСТРАЦІЇ ІНФОРМАЦІЇ

вул. М.Шпака, 2, м. Київ, 03113; Тел. (44) 456-83-89, (44) 454-21-51; Факс: (44) 456-33-18;

E-mail: ipri@ipri.kiev.ua; <http://www.ipri.kiev.ua>; Код ЄДРПОУ 03771755

13.01.2021 № 104-3/152
 На № _____

А К Т

про впровадження результатів дисертаційної роботи
 Єфремова Костянтина Вікторовича на тему «Платформа підтримки міждисциплінарних досліджень в Світовій системі даних», поданої на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – Інформаційні технології

Платформа підтримки міждисциплінарних досліджень в Світовій системі даних, що розроблена в дисертаційній роботі Єфремова К. В., використана при створенні інформаційних систем Інституту проблем реєстрації інформації Національної академії наук України. Впровадження положень дисертаційної роботи дало такі результати:

1. Використання запропонованих в дисертаційній роботі архітектурних принципів організації системи інтелектуальної обробки даних, моделей і методів та розроблених на їх основі інструментів інтеграції застосунків і джерел даних дозволило створювати застосунки для інтелектуального аналізу даних та прогнозування поведінки складних систем різної природи.
2. Використання запропонованих в дисертаційній роботі, моделей і методів та розроблених на їх основі інструментів генерування проблемно-орієнтованих застосунків для аналітичного оброблення міждисциплінарних даних дозволило скоротити час проектування і реалізації застосунків на 20%.
3. Впровадження платформи в роботу наукових підрозділів, пов'язаних з міждисциплінарними дослідженнями, дозволило підвищити обсяги наданих користувачам інформаційних ресурсів на 50%.
4. Запропоновані в роботі моделі агента та міжагентної взаємодії, моделі і методи побудови зв'язаної структури даних із онтологій та підключень до джерел даних використано при розв'язанні проблем моделювання предметних областей при формуванні баз знань і забезпеченні семантичного пошуку.
5. Експлуатаційні випробування інформаційних систем, побудованих на основі платформи підтримки міждисциплінарних досліджень в Світовій системі даних, запропонованої в дисертаційній роботі, продемонстрували їх гнучкість та адаптивність.

Директор,
 академік НАН України, д.т.н., проф.



В. В. Петров



SCIENCE · EDUCATION · BUSINESS

Науковий парк “Київська політехніка”

№ 21-002/1-НП
27 січня 2021 р.

АКТ

про впровадження результатів дисертаційної роботи Єфремова Костянтина Вікторовича на тему «Платформа підтримки міждисциплінарних досліджень в Світовій системі даних», поданої на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – Інформаційні технології

Корпорація «Науковий парк «Київська політехніка» активно працює в сфері розвитку та підтримки науково-технічної та інноваційної діяльності, ефективного та раціонального використання наявного наукового потенціалу, матеріально-технічної бази для комерціалізації результатів наукових досліджень і їх впровадження на вітчизняному та закордонному ринках. Для реалізації статутних задач Корпорація організовує співпрацю наукових і дослідницьких організацій, наукових груп та високотехнологічних компаній для створення наукоємної продукції, в тому числі в сфері національної безпеки і оборони держави, шляхом координації виконання спільних інноваційних проектів. Це потребує інструментів для підтримки міждисциплінарних досліджень, інтелектуальної обробки великих обсягів даних, отриманих з джерел різної природи, узгодження даних, швидкого створення і надання інформаційних ресурсів у зручній для опрацювання формі.

Розроблені в рамках дисертаційної роботи Єфремова К. В. математичні моделі та методи інтеграції застосунків і джерел даних, узгодження даних, впровадження проблемно-орієнтованого апарату інтелектуального аналізу даних та швидкого створення інформаційних ресурсів, а також створені на їх основі інструментальні засоби платформи підтримки міждисциплінарних досліджень впроваджені і успішно використовуються в роботі Корпорації «Науковий парк «Київська політехніка» та мережі організацій-партнерів.

Генеральний директор



Віктор КАМАСЬВ



УКРАЇНА

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

03056, м. Київ, просп. Перемоги, 37; тел. (+38 044) 204-82-82; тел./факс (+38 044) 204-97-88
<http://www.kpi.ua> e-mail: mail@kpi.ua ЕДРПОУ 02070921

25.01.2021 № 01-24/21
на № _____ « _____ » _____ 20__

АКТ

про впровадження результатів дисертаційної роботи
Єфремова Костянтина Вікторовича на тему «Платформа підтримки
міждисциплінарних досліджень в Світовій системі даних», поданої на
здобуття наукового ступеня кандидата технічних наук за спеціальністю
05.13.06 – Інформаційні технології

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» цим актом підтверджує, що впровадження положень дисертаційної роботи Єфремова К.В. дозволило отримати такі результати:

1. Розроблені методичні та програмні засоби використовуються на кафедрі математичних методів системного аналізу та кафедрі автоматики та управління в технічних системах КПІ ім. Ігоря Сікорського при підготовці фахівців в галузі інтелектуального аналізу даних та інформаційних технологій, зокрема в навчальних дисциплінах «Інфраструктура інформаційних технологій», «Засоби розробки прикладних і системних програм», «Розроблення WEB-застосувань», «Бази даних», «Проектування інформаційних систем», «Методи та засоби надання інформаційних сервісів», «Інтелектуальний аналіз даних».

Впровадження результатів дисертаційної роботи Єфремова К. В. дає можливість ознайомити студентів з новими підходами і технологіями створення інформаційних систем для організації і підтримки дослідницьких інфраструктур.

2. Платформу підтримки міждисциплінарних досліджень в Світовій системі даних, розроблену в дисертаційній роботі Єфремова К. В., використано при створенні апаратно-програмного комплексу Інформаційно-аналітичного ситуаційного центру КПІ ім. Ігоря Сікорського для інтеграції програмних ресурсів (джерел даних, сервісів) в рамках єдиного відкритого гетерогенного розподіленого інформаційного середовища та реалізації інструментів підтримки робочих процесів колективного напрацювання управлінських рішень на основі інтелектуальної обробки даних великих обсягів.

3. В рамках інтеграційної підсистеми Інформаційно-аналітичного ситуаційного центру КПІ ім. Ігоря Сікорського при розв'язанні проблем інтеграції даних та застосунків використовуються запропоновані в дисертаційній роботі логічні моделі опису застосунків та джерел даних, визначення семантики предметної області, перетворення даних, агента та міжагентної взаємодії, методи та алгоритми виведення, відновлення дерева виконання запиту, метод узгодження даних з врахуванням інформаційних втрат, модель і метод генерування проблемно-орієнтованих застосунків на основі інтеграції мікросервісів та методичні рекомендації щодо створення та налаштування таких застосунків.

4. З використанням інструментальних засобів платформи підтримки міждисциплінарних досліджень в Світовій системі даних в Інформаційно-аналітичному ситуаційному центрі КПІ ім. Ігоря Сікорського розроблено і підтримується низка проблемно-орієнтованих застосунків, спрямованих на упередження і подолання наслідків кризових явищ різної природи та забезпечення сталого розвитку суспільства в умовах негативного впливу множинних загроз економіці і національній безпеці України, зокрема спеціалізований застосунок «Pro Et Contra», призначений для кількісної оцінки параметрів та характеристик суспільних перетворень на основі методів сентимент-аналізу та інтелектуального аналізу текстових повідомлень з медіа-ресурсів та соціальних мереж, експертних оцінок.

Проректор з навчальної роботи



Анатолій МЕЛЬНИЧЕНКО



УКРАЇНА

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
«СВІТОВИЙ ЦЕНТР ДАНИХ З ГЕОІНФОРМАТИКИ ТА СТАЛОГО РОЗВИТКУ»

03056, м. Київ, пр. Перемоги, 37; тел. (+38 044) 204 8014 тел./факс (+38 044) 204 8153
web: <http://wdc.org.ua>, e-mail: mail@wdc.org.ua

02.02.2021 № 0302/21
на № _____ від _____

АКТ

про впровадження результатів дисертаційної роботи
Єфремова Костянтина Вікторовича на тему «Платформа підтримки
міждисциплінарних досліджень в Світовій системі даних», поданої на здобуття
наукового ступеня кандидата технічних наук за спеціальністю
05.13.06 – Інформаційні технології

Результати дисертаційної роботи Єфремова К. В. використано при створенні інтегрованої он-лайн платформи «Advanced Analytics» для сценарного моделювання кризових і безпекових явищ та вивчення їх впливу на економіку і суспільство, забезпечивши глибшу інтеграцію ННК «Світовий центр даних з геоінформатики та сталого розвитку» у IT-середовище Світової системи даних та зростання обсягів і масштабів виконуваних міждисциплінарних досліджень.

1. Впровадження запропонованої в дисертаційній роботі платформи дозволило вирішити задачу організації, планування і здійснення підтримки міждисциплінарних досліджень в ННК «Світовий центр даних з геоінформатики та сталого розвитку» (СЦД-Україна), підвищити обсяги наданих користувачам інформаційних ресурсів на 50%.
2. Впровадження інструментальних засобів платформи підтримки міждисциплінарних досліджень в Світовій системі даних забезпечило можливість реалізації мікросервісів, сценаріїв обробки даних та проблемно-орієнтованих застосунків для вирішення низки прикладних задач, зокрема: аналізу і сценарного моделювання сталого розвитку

- територій; моделювання, кількісного оцінювання та прогнозування впливу загроз виникнення конфліктів на стабільність розвитку суспільства; системного дослідження та прогнозного моделювання циклічного виникнення і поширення пандемій інфекційних хвороб (з урахуванням пандемії COVID-19) в Україні на довгостроковому (декілька років), середньостроковому (декілька місяців) і короткостроковому (5-7 днів) часових горизонтах.
3. Використання запропонованих в дисертаційній роботі моделей і методів та розроблених на їх основі інструментів генерування проблемно-орієнтованих застосунків для аналітичного оброблення міждисциплінарних даних дозволило скоротити час проектування і реалізації тематичних застосунків на 27 %.
 4. Запропоновані в дисертаційній роботі архітектурні принципи організації системи інтелектуальної обробки даних, моделі та методи інтеграції джерел даних і застосунків, методи перетворення та узгодження даних з врахуванням інформаційних втрат, а також реалізовані на їх основі інструментальні засоби було використано і впроваджено в рамках низки держбюджетних і договірних науково-дослідних робіт та проектів, виконуваних СЦД-Україна, серед яких: проєкт «Розробка інформаційних технологій для моделювання, кількісного оцінювання та прогнозування впливу загроз виникнення конфліктів та розповсюдження зброї на стабільність розвитку суспільства в регіональному та глобальному масштабах» (2016-2017 рр, проєкт №6166, Український науково-технологічний центр (STCU); НДР на замовлення підприємств і організацій України «Форсайт та побудова стратегії соціально-економічного розвитку України на середньостроковому (до 2020 року) і довгостроковому (до 2030 року) часових горизонтах» (2016 р., Кооператив «Перший національний аграрний кооператив», ГО «Фундація Аграрна наддержжава»); НДР «Розробка он-лайн платформи аналізу і сценарного планування сталого розвитку регіонів України в контексті якості та безпеки життя людей» (2017-2019 рр., РК №0117U002476, МОН України); НДР «Інтегрована платформа для оцінювання та сценарного планування сталого розвитку об'єднаних територіальних громад в ході проведення адміністративно-територіальної реформи в Україні» (2020-2022 рр., РК 0120U102060, МОН України) та інші.

Заступник директора,
к.ф.-м.н., доцент



Іван ПИШНОГРАЄВ